

New York State Testing Program 2017: English Language Arts and Mathematics Grades 3–8



Technical Report

**Questar Assessment Inc.
2017**

Developed and published under contract with the New York State Education Department by Questar Assessment Inc., 5550 Upper 147th Street West, Apple Valley, MN 55124. Copyright © 2017 by the New York State Education Department.

Permission is hereby granted for New York State school administrators and educators to reproduce these materials, located online at <http://www.p12.nysed.gov/assessment/reports/>, in the quantities necessary for educational researchers' use, but not for sale, provided copyright notices are retained as they appear in these publications.

Table of Contents

Section 1: Introduction and Overview	1
1.1. Introduction.....	1
1.2. Test Purpose.....	1
1.3. Expected Participants.....	1
1.4. Test Use and Decisions Based on Assessment	1
1.4.1. Scale Scores.....	1
1.4.2. Statewide Percentile Ranks	2
1.4.3. Performance Level Cut Scores and Classification	2
1.4.4. Subscores.....	2
1.5. Testing Accommodations	4
1.6. Test Transcriptions.....	4
1.7. Test Translations.....	4
Section 2: Test Design and Development.....	6
2.1. Test Descriptions	6
2.1.1. ELA Tests.....	6
2.1.2. Mathematics Tests.....	6
2.2. Test Configuration	7
2.2.1. Test Book Design	7
2.2.2. Embedded Field-Test Items.....	7
2.3. New York State Educators’ Involvement in Test Development.....	8
2.4. Test Blueprints	8
2.5. Passage Selection and Item Criteria Documents	9
2.5.1. Principles of Universal Design.....	10
2.6. Passage Finding	11
2.7. Item Development.....	11
2.8. Educator Item Review.....	12
2.9. Field-Testing.....	13
2.10. Rangefinding.....	14
2.11. Item Selection and Test Creation (Criteria and Process).....	14
2.12. Educator Form Construction.....	15
2.13. Test Form Production	16
2.14. Final Eyes Committees	16
2.15. Proficiency and Performance Standards	16
Section 3: Validity	18
3.1. Content Validity.....	18
3.2. Construct (Internal Structure) Validity	19
3.2.1. Internal Consistency	19
3.2.2. Unidimensionality	19
3.2.3. Detection of Bias	23
Section 4: Test Administration and Scoring.....	24

4.1. Test Administration	24
4.2. Scoring Procedures of Operational Tests	24
4.3. Scoring Models	25
4.4. Scoring of Constructed-Response Items	25
4.5. Scorer Qualifications and Training	26
4.6. Quality Control Process	26
Section 5: Operational Test Data Collection and Classical Analysis	27
5.1. Data Collection	27
5.2. Data Processing	27
5.2.1. Sampling Down for Representativeness	27
5.3. Classical Analysis and Calibration Sample Characteristics	32
5.4. Classical Data Analysis	42
5.4.1. Item Difficulty and Point Biserial Correlation Coefficients	42
5.4.2. Omit Rates	43
5.4.3. Differential Item Functioning (DIF)	44
Section 6: IRT Calibration and Linking	47
6.1. IRT Models and Rationale for Use	47
6.2. Calibration Sample	48
6.2.1. Calibration Process	53
6.3. Item-Model Fit	54
6.4. Local Independence	56
6.5. Linking and Scaling	57
6.6. Anchor Set Evaluation	59
6.7. Test Characteristic Curves	61
6.8. Scoring Procedure	73
6.8.1. Raw Score-to-Scale Score and SEM Conversion Tables	74
Section 7: Reliability and Standard Error of Measurement	76
7.1. Test Reliability	76
7.1.1. Test Statistics and Reliability for Total Test	76
7.1.2. Reliability of MC Items	78
7.1.3. Reliability of CR Items	78
7.1.4. Test Reliability for Reporting Categories	79
7.2. Standard Error of Measurement (SEM)	89
7.3. Performance Level Classification Consistency and Accuracy	90
7.3.1. Consistency	90
7.3.2. Accuracy	92
Section 8: Summary of Operational Test Results	94
8.1. Scale Score Distribution Summary	94
8.1.1. ELA Scale Score and Subscore Distributions	94
8.1.1.1. ELA Grade 3	95
8.1.1.2. ELA Grade 4	96
8.1.1.3. ELA Grade 5	97

8.1.1.4. ELA Grade 6	98
8.1.1.5. ELA Grade 7	99
8.1.1.6. ELA Grade 8	100
8.1.2. Mathematics Scale Score Distributions	102
8.1.2.1. Mathematics Grade 3.....	103
8.1.2.2. Mathematics Grade 4.....	104
8.1.2.3. Mathematics Grade 5.....	105
8.1.2.4. Mathematics Grade 6.....	107
8.1.2.5. Mathematics Grade 7.....	108
8.1.2.6. Mathematics Grade 8.....	109
8.2. Performance Level Distribution Summary	111
8.2.1. ELA Test Performance Level Distributions	111
8.2.1.1. ELA Grade 3	112
8.2.1.2. ELA Grade 4	113
8.2.1.3. ELA Grade 5	114
8.2.1.4. ELA Grade 6	115
8.2.1.5. ELA Grade 7	116
8.2.1.6. ELA Grade 8	117
8.2.2. Mathematics Test Performance Level Distributions	118
8.2.2.1. Mathematics Grade 3.....	119
8.2.2.2. Mathematics Grade 4.....	120
8.2.2.3. Mathematics Grade 5.....	121
8.2.2.4. Mathematics Grade 6.....	123
8.2.2.5. Mathematics Grade 7.....	124
8.2.2.6. Mathematics Grade 8.....	125
Section 9: References.....	128
Appendix A: ELA and Mathematics Test Configurations.....	131
Appendix B: ELA and Mathematics Test Blueprints	135
Appendix C: Passage Selection Guidelines for Assessing ELA.....	137
Appendix D: Universal Design Item Checklist	138
Appendix E: Criteria for Item Acceptability	141
Appendix F: Psychometric Guidelines for Operational Item Selection.....	143
Appendix G: Operational Item Maps.....	144
Appendix H: ELA Short-Response Rubric.....	160
Appendix I: ELA Extended-Response Rubric.....	161
Appendix J: Mathematics Short-Response Rubric	164
Appendix K: Mathematics Extended-Response Rubric	165
Appendix L: Factor Analysis Results for Select Subgroups.....	166
Appendix M: Classical Test Theory Statistics.....	183
Appendix N: Items Flagged for DIF	199
Appendix O: IRT Statistics	204
Appendix P: Derivation and Estimation of Classification Consistency and Accuracy	235
Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables.....	238
Appendix R: Study of Operational Test Mode Comparability	293

Appendix S: Memo on Operational Test Mode Comparability.....	338
--	-----

List of Tables

Table 1.1. ELA Subscore Categories and Total Possible Score Points	3
Table 1.2. Mathematics Subscore Categories and Total Possible Score Points	3
Table 2.1. Summary of Unique 2016 Field Test Items	13
Table 3.1. ELA Tests Factor Analysis	21
Table 3.2. Mathematics Tests Factor Analysis	22
Table 5.1. ELA Grade 3 Data Cleaning	28
Table 5.2. ELA Grade 4 Data Cleaning	28
Table 5.3. ELA Grade 5 Data Cleaning	29
Table 5.4. ELA Grade 6 Data Cleaning	29
Table 5.5. ELA Grade 7 Data Cleaning	29
Table 5.6. ELA Grade 8 Data Cleaning	30
Table 5.7. Mathematics Grade 3 Data Cleaning	30
Table 5.8. Mathematics Grade 4 Data Cleaning	30
Table 5.9. Mathematics Grade 5 Data Cleaning	31
Table 5.10. Mathematics Grade 6 Data Cleaning	31
Table 5.11. Mathematics Grade 7 Data Cleaning	31
Table 5.12. Mathematics Grade 8 Data Cleaning	32
Table 5.13. ELA Grade 3 Sample Characteristics	32
Table 5.14. ELA Grade 4 Sample Characteristics	33
Table 5.15. ELA Grade 5 Sample Characteristics	34
Table 5.16. ELA Grade 6 Sample Characteristics	34
Table 5.17. ELA Grade 7 Sample Characteristics	35
Table 5.18. ELA Grade 8 Sample Characteristics	36
Table 5.19. Mathematics Grade 3 Sample Characteristics	36
Table 5.20. Mathematics Grade 4 Sample Characteristics	37
Table 5.21. Mathematics Grade 5 Sample Characteristics	38
Table 5.22. Mathematics Grade 6 Sample Characteristics	40
Table 5.23. Mathematics Grade 7 Sample Characteristics	40
Table 5.24. Mathematics Grade 8 Sample Characteristics	41
Table 5.25. ELA Classical DIF Sample N-Counts	45
Table 5.26. Mathematics Classical DIF Sample N-Counts	45
Table 5.27. ELA Items Flagged for DIF	46
Table 5.28. Mathematics Items Flagged for DIF	46
Table 6.1. ELA Grades 3 and 4 Demographic Statistics	48
Table 6.2. ELA Grades 5 and 6 Demographic Statistics	49
Table 6.3. ELA Grades 7 and 8 Demographic Statistics	50
Table 6.4. Mathematics Grades 3 and 4 Demographic Statistics	51
Table 6.5. Mathematics Grades 5 and 6 Demographic Statistics	52
Table 6.6. Mathematics Grades 7 and 8 Demographic Statistics	52

Table 6.7. ELA Calibration Results	54
Table 6.8. Mathematics Calibration Results	54
Table 6.9. ELA Linking Coefficients	57
Table 6.10. Mathematics Linking Coefficients	57
Table 6.11. ELA Scaling Coefficients	59
Table 6.12. Mathematics Scaling Coefficients	59
Table 6.13. PBT ELA Performance-Level Scale Score Ranges	75
Table 6.14. PBT Mathematics Performance-Level Scale Score Ranges	75
Table 6.15. CBT ELA Performance-Level Scale Score Ranges ...	Error! Bookmark not defined.
Table 6.16. CBT Mathematics Performance-Level Scale Score Ranges.....	Error! Bookmark not defined.
Table 7.1. ELA Test Form Statistics	77
Table 7.2. ELA Test Reliability and Standard Error of Measurement	77
Table 7.3. Mathematics Test Form Statistics	77
Table 7.4. Mathematics Test Reliability and Standard Error of Measurement	77
Table 7.5. ELA MC Item Reliability and Standard Error of Measurement	78
Table 7.6. Mathematics MC Item Reliability and Standard Error of Measurement	78
Table 7.7. ELA CR Item Reliability and Standard Error of Measurement	79
Table 7.8. Mathematics CR Item Reliability and Standard Error of Measurement	79
Table 7.9. ELA Grade 3 Test Reliability by Subgroup	80
Table 7.10. ELA Grade 4 Test Reliability by Subgroup	80
Table 7.11. ELA Grade 5 Test Reliability by Subgroup	81
Table 7.12. ELA Grade 6 Test Reliability by Subgroup	82
Table 7.13. ELA Grade 7 Test Reliability by Subgroup	82
Table 7.14. ELA Grade 8 Test Reliability by Subgroup	83
Table 7.15. Mathematics Grade 3 Test Reliability by Subgroup	84
Table 7.16. Mathematics Grade 4 Test Reliability by Subgroup	85
Table 7.17. Mathematics Grade 5 Test Reliability by Subgroup	86
Table 7.18. Mathematics Grade 6 Test Reliability by Subgroup	87
Table 7.19. Mathematics Grade 7 Test Reliability by Subgroup	88
Table 7.20. Mathematics Grade 8 Test Reliability by Subgroup	89
Table 7.21. Decision Consistency (All Cuts)*	91
Table 7.22. Decision Consistency (Level III Cut)*	92
Table 7.23. Decision Agreement (Accuracy) Estimates*	93
Table 8.1. ELA Scale Score Distribution Summary	94
Table 8.2. ELA Subscore Summary	95
Table 8.3. ELA Grade 3 Scale Score Distribution by Subgroup	95
Table 8.4. ELA Grade 4 Scale Score Distribution by Subgroup	97
Table 8.5. ELA Grade 5 Scale Score Distribution by Subgroup	98
Table 8.6. ELA Grade 6 Scale Score Distribution by Subgroup	99
Table 8.7. ELA Grade 7 Scale Score Distribution by Subgroup	100

Table 8.8. ELA Grade 8 Scale Score Distribution by Subgroup	101
Table 8.9. Mathematics Scale Score Distribution Summary	102
Table 8.10. Mathematics Subscore Summary.....	102
Table 8.11. Mathematics Grade 3 Scale Score Distribution by Subgroup	103
Table 8.12. Mathematics Grade 4 Scale Score Distribution by Subgroup	105
Table 8.13. Mathematics Grade 5 Scale Score Distribution by Subgroup	106
Table 8.14. Mathematics Grade 6 Scale Score Distribution by Subgroup	107
Table 8.15. Mathematics Grade 7 Scale Score Distribution by Subgroup	109
Table 8.16. Mathematics Grade 8 Scale Score Distribution by Subgroup	110
Table 8.17. ELA Test Performance Level Distributions	111
Table 8.18. ELA Grade 3 Performance Level Distribution by Subgroup	112
Table 8.19. ELA Grade 4 Performance Level Distribution by Subgroup	113
Table 8.20. ELA Grade 5 Performance Level Distribution by Subgroup	114
Table 8.21. ELA Grade 6 Performance Level Distribution by Subgroup	115
Table 8.22. ELA Grade 7 Performance Level Distribution by Subgroup	116
Table 8.23. ELA Grade 8 Performance Level Distribution by Subgroup	118
Table 8.24. Mathematics Test Performance Level Distributions	119
Table 8.25. Mathematics Grade 3 Performance Level Distribution by Subgroup.....	119
Table 8.26. Mathematics Grade 4 Performance Level Distribution by Subgroup.....	121
Table 8.27. Mathematics Grade 5 Performance Level Distribution by Subgroup.....	122
Table 8.28. Mathematics Grade 6 Performance Level Distribution by Subgroup.....	123
Table 8.29. Mathematics Grade 7 Performance Level Distribution by Subgroup.....	124
Table 8.30. Mathematics Grade 8 Performance Level Distribution by Subgroup.....	126
Table A1. ELA Test Configuration	131
Table A2. Mathematics Test Configuration.....	131
Table A3. ELA Estimated Time on Task by Book.....	132
Table A4. Mathematics Estimated Time on Task by Book	133
Table B1. ELA Test Blueprint	135
Table B2. Mathematics Test Blueprint	135
Table G1. ELA Grade 3 Operational Item Map	144
Table G2. ELA Grade 4 Operational Item Map	145
Table G3. ELA Grade 5 Operational Item Map	146
Table G4. ELA Grade 6 Operational Item Map	147
Table G5. ELA Grade 7 Operational Item Map	148
Table G6. ELA Grade 8 Operational Item Map	149
Table G7. Mathematics Grade 3 Operational Item Map.....	151
Table G8. Mathematics Grade 4 Operational Item Map.....	152
Table G9. Mathematics Grade 5 Operational Item Map.....	153
Table G10. Mathematics Grade 6 Operational Item Map.....	155
Table G11. Mathematics Grade 7 Operational Item Map.....	156
Table G12. Mathematics Grade 8 Operational Item Map.....	158

Table L1. ELA Grade 3 Test Factor Analysis by Subgroup.....	166
Table L2. ELA Grade 4 Test Factor Analysis by Subgroup.....	167
Table L3. ELA Grade 5 Test Factor Analysis by Subgroup.....	168
Table L4. ELA Grade 6 Test Factor Analysis by Subgroup.....	170
Table L5. ELA Grade 7 Test Factor Analysis by Subgroup.....	171
Table L6. ELA Grade 8 Test Factor Analysis by Subgroup.....	173
Table L7. Mathematics Grade 3 Test Factor Analysis by Subgroup.....	174
Table L8. Mathematics Grade 4 Test Factor Analysis by Subgroup.....	175
Table L9. Mathematics Grade 5 Test Factor Analysis by Subgroup.....	176
Table L10. Mathematics Grade 6 Test Factor Analysis by Subgroup.....	177
Table L11. Mathematics Grade 7 Test Factor Analysis by Subgroup.....	179
Table L12. Mathematics Grade 8 Test Factor Analysis by Subgroup.....	180
Table M1. ELA Grade 3 Classical Item Analysis.....	183
Table M2. ELA Grade 4 Classical Item Analysis.....	184
Table M3. ELA Grade 5 Classical Item Analysis.....	185
Table M4. ELA Grade 6 Classical Item Analysis.....	186
Table M5. ELA Grade 7 Classical Item Analysis.....	187
Table M6. ELA Grade 8 Classical Item Analysis.....	188
Table M7. Mathematics Grade 3 Classical Item Analysis.....	190
Table M8. Mathematics Grade 4 Classical Item Analysis.....	191
Table M9. Mathematics Grade 5 Classical Item Analysis.....	192
Table M10. Mathematics Grade 6 Classical Item Analysis.....	194
Table M11. Mathematics Grade 7 Classical Item Analysis.....	195
Table M12. Mathematics Grade 8 Classical Item Analysis.....	197
Table N1. ELA MC Item Classical DIF Flags.....	199
Table N2. ELA CR Item Classical DIF Flags.....	201
Table N3. Mathematics MC Item Classical DIF Flags.....	202
Table N4. Mathematics CR Item Classical DIF Flags.....	203
Table O1. ELA Grade 3 Item Fit Statistics.....	204
Table O2. ELA Grade 4 Item Fit Statistics.....	205
Table O3. ELA Grade 5 Item Fit Statistics.....	206
Table O4. ELA Grade 6 Item Fit Statistics.....	207
Table O5. ELA Grade 7 Item Fit Statistics.....	208
Table O6. ELA Grade 8 Item Fit Statistics.....	209
Table O7. Mathematics Grade 3 Item Fit Statistics.....	211
Table O8. Mathematics Grade 4 Item Fit Statistics.....	212
Table O9. Mathematics Grade 5 Item Fit Statistics.....	213
Table O10. Mathematics Grade 6 Item Fit Statistics.....	215
Table O11. Mathematics Grade 7 Item Fit Statistics.....	216
Table O12. Mathematics Grade 8 Item Fit Statistics.....	218
Table O13. ELA Grade 3 OP Item Parameter Estimates.....	219

Table O14. ELA Grade 4 OP Item Parameter Estimates.....	220
Table O15. ELA Grade 5 OP Item Parameter Estimates.....	221
Table O16. ELA Grade 6 OP Item Parameter Estimates.....	222
Table O17. ELA Grade 7 OP Item Parameter Estimates.....	224
Table O18. ELA Grade 8 OP Item Parameter Estimates.....	225
Table O19. Mathematics Grade 3 OP Item Parameter Estimates.....	226
Table O20. Mathematics Grade 4 OP Item Parameter Estimates.....	227
Table O21. Mathematics Grade 5 OP Item Parameter Estimates.....	229
Table O22. Mathematics Grade 6 OP Item Parameter Estimates.....	230
Table O23. Mathematics Grade 7 OP Item Parameter Estimates.....	232
Table O24. Mathematics Grade 8 OP Item Parameter Estimates.....	233
Table Q1. PBT ELA Grade 3 RSSS Table.....	238
Table Q2. PBT ELA Grade 4 RSSS Table.....	238
Table Q3. PBT ELA Grade 5 RSSS Table.....	239
Table Q4. PBT ELA Grade 6 RSSS Table.....	240
Table Q5. PBT ELA Grade 7 RSSS Table.....	241
Table Q6. PBT ELA Grade 8 RSSS Table.....	242
Table Q7. PBT Mathematics Grade 3 RSSS Table.....	243
Table Q8. PBT Mathematics Grade 4 RSSS Table.....	244
Table Q9. PBT Mathematics Grade 5 RSSS Table.....	245
Table Q10. PBT Mathematics Grade 6 RSSS Table.....	246
Table Q11. PBT Mathematics Grade 7 RSSS Table.....	247
Table Q12. PBT Mathematics Grade 8 RSSS Table.....	248
Table Q13. CBT ELA Grade 3 RSSS Table.....	249
Table Q14. CBT ELA Grade 4 RSSS Table.....	250
Table Q15. CBT ELA Grade 5 RSSS Table.....	251
Table Q16. CBT ELA Grade 6 RSSS Table.....	252
Table Q17. CBT ELA Grade 7 RSSS Table.....	253
Table Q18. CBT ELA Grade 8 RSSS Table.....	254
Table Q19. CBT Mathematics Grade 3 RSSS Table.....	255
Table Q20. CBT Mathematics Grade 4 RSSS Table.....	256
Table Q21. CBT Mathematics Grade 5 RSSS Table.....	257
Table Q22. CBT Mathematics Grade 6 RSSS Table.....	258
Table Q23. CBT Mathematics Grade 7 RSSS Table.....	259
Table Q24. CBT Mathematics Grade 8 RSSS Table.....	260
Table Q25. PBT ELA Grade 3 Scale Score Frequency Distribution.....	261
Table Q26. PBT ELA Grade 4 Scale Score Frequency Distribution.....	263
Table Q27. PBT ELA Grade 5 Scale Score Frequency Distribution.....	265
Table Q28. PBT ELA Grade 6 Scale Score Frequency Distribution.....	268
Table Q29. PBT ELA Grade 7 Scale Score Frequency Distribution.....	271
Table Q30. PBT ELA Grade 8 Scale Score Frequency Distribution.....	274

Table Q31. PBT Mathematics Grade 3 Scale Score Frequency Distribution	276
Table Q32. PBT Mathematics Grade 4 Scale Score Frequency Distribution	278
Table Q33. PBT Mathematics Grade 5 Scale Score Frequency Distribution	281
Table Q34. PBT Mathematics Grade 6 Scale Score Frequency Distribution	284
Table Q35. PBT Mathematics Grade 7 Scale Score Frequency Distribution	287
Table Q36. PBT Mathematics Grade 8 Scale Score Frequency Distribution	289
Table R.1.1. Unique Items Administered in Both CBT and PBT Modes: ELA	293
Table R.1.2. Unique Items Administered in Both CBT and PBT Modes: Mathematics	293
Table R.2.1. Sample Sizes Before Matching by Mode: ELA	294
Table R.2.2. Sample Sizes Before Matching by Mode: Mathematics	294
Table R.3.1. Covariate Balance Before and After Matching: ELA Grade 4	298
Table R.3.2. Covariate Balance Before and After Matching: ELA Grade 5	300
Table R.3.3. Covariate Balance Before and After Matching: ELA Grade 6	302
Table R.3.4. Covariate Balance Before and After Matching: ELA Grade 7	304
Table R.3.5. Covariate Balance Before and After Matching: ELA Grade 8	306
Table R.3.6. Covariate Balance Before and After Matching: Mathematics Grade 4	308
Table R.3.7. Covariate Balance Before and After Matching: Mathematics Grade 5	310
Table R.3.8. Covariate Balance Before and After Matching: Mathematics Grade 6	312
Table R.3.9. Covariate Balance Before and After Matching: Mathematics Grade 7	314
Table R.3.10. Covariate Balance Before and After Matching: Mathematics Grade 8	316
Table R.3.11. Test Performance After Matching: ELA	318
Table R.3.12. Test Performance After Matching: Mathematics	319
Table R.4.1. CBT Scale Score Adjustments	321
Table R.A.1. Propensity Score Model: ELA Grade 4	324
Table R.A.2. Propensity Score Model: ELA Grade 5	325
Table R.A.3. Propensity Score Model: ELA Grade 6	326
Table R.A.4. Propensity Score Model: ELA Grade 7	327
Table R.A.5. Propensity Score Model: ELA Grade 8	328
Table R.A.6. Propensity Score Model: Mathematics Grade 4	329
Table R.A.7. Propensity Score Model: Mathematics Grade 5	330
Table R.A.8. Propensity Score Model: Mathematics Grade 6	331
Table R.A.9. Propensity Score Model: Mathematics Grade 7	332
Table R.A.10. Propensity Score Model: Mathematics Grade 8	333

List of Figures

Figure 6.1. ELA Grade 3 TCC.....	61
Figure 6.2. ELA Grade 3 CSEM Curve.....	62
Figure 6.3. ELA Grade 4 TCC.....	62
Figure 6.4. ELA Grade 4 CSEM Curve.....	63
Figure 6.5. ELA Grade 5 TCC.....	63
Figure 6.6. ELA Grade 5 CSEM Curve.....	64
Figure 6.7. ELA Grade 6 TCC.....	64
Figure 6.8. ELA Grade 6 CSEM Curve.....	65
Figure 6.9. ELA Grade 7 TCC.....	65
Figure 6.10. ELA Grade 7 CSEM Curve.....	66
Figure 6.11. ELA Grade 8 TCC.....	66
Figure 6.12. ELA Grade 8 CSEM Curve.....	67
Figure 6.13. Mathematics Grade 3 TCC.....	67
Figure 6.14. Mathematics Grade 3 CSEM Curve.....	68
Figure 6.15. Mathematics Grade 4 TCC.....	68
Figure 6.16. Mathematics Grade 4 CSEM Curve.....	69
Figure 6.17. Mathematics Grade 5 TCC.....	69
Figure 6.18. Mathematics Grade 5 CSEM Curve.....	70
Figure 6.19. Mathematics Grade 6 TCC.....	70
Figure 6.20. Mathematics Grade 6 CSEM Curve.....	71
Figure 6.21. Mathematics Grade 7 TCC.....	71
Figure 6.22. Mathematics Grade 7 CSEM Curve.....	72
Figure 6.23. Mathematics Grade 8 TCC.....	72
Figure 6.24. Mathematics Grade 8 CSEM Curve.....	73
Figure R.3.1. Key Covariate Balance Before and After Matching: ELA Grade 4	299
Figure R.3.2. Key Covariate Balance Before and After Matching: ELA Grade 5	301
Figure R.3.3. Key Covariate Balance Before and After Matching: ELA Grade 6	303
Figure R.3.4. Key Covariate Balance Before and After Matching: ELA Grade 7	305
Figure R.3.5. Key Covariate Balance Before and After Matching: ELA Grade 8	307
Figure R.3.6. Key Covariate Balance Before and After Matching: Mathematics Grade 4	309
Figure R.3.7. Key Covariate Balance Before and After Matching: Mathematics Grade 5	311
Figure R.3.8. Key Covariate Balance Before and After Matching: Mathematics Grade 6	313
Figure R.3.9. Key Covariate Balance Before and After Matching: Mathematics Grade 7	315
Figure R.3.10. Key Covariate Balance Before and After Matching: Mathematics Grade 8	317
Figure R.A.1. Distribution of Propensity Scores Before and After Matching: ELA.....	334
Figure R.A.2. Distribution of Propensity Scores Before and After Matching: Mathematics	335
Figure R.A.3. Distribution of Raw Score Differences After Matching: ELA	336
Figure R.A.4. Distribution of Raw Score Differences After Matching: Mathematics	337

Section 1: Introduction and Overview

1.1. Introduction

This technical report provides detailed information regarding the technical, statistical, and measurement attributes of the New York State Testing Program (NYSTP) for the Grades 3–8 English Language Arts (ELA) and Mathematics 2017 Operational Tests. This report includes information about test content and test development, item (i.e., individual test question) and test statistics, validity and reliability, differential item functioning (DIF) studies, test administration, scoring, linking, scaling, and student performance.

1.2. Test Purpose

The 2017 Grades 3–8 ELA and Mathematics NYSTP has been designed to measure student knowledge and skills as defined by grade-level New York State Learning Standards (CCLS) in ELA and Mathematics. The tests are designed to allow the classification of student proficiency into four performance levels (Level I, Level II, Level III, and Level IV). Likewise, the test provides students at each of these performance levels opportunities to demonstrate their knowledge and skills in the CCLS. Details about the content standards for ELA and Mathematics are described in Section 2.4: Test Blueprints.

1.3. Expected Participants

Students in New York State public school grades 3, 4, 5, 6, 7, and 8 (and ungraded students of equivalent chronological ages) are the expected participants for the Grades 3–8 NYSTP. Religious and independent schools may participate in the testing program, but their participation is not mandatory. In 2017, some religious and independent schools participated in the testing program across all grade levels. These schools were included in the data analyses. Public school students were required to take all State assessments administered at their grade level, except for a very small percentage of students with severe cognitive disabilities who took the New York State Alternate Assessment (NYSAA). For more detail on this exemption, please refer to the *NYSTP Grades 3–8 English Language Arts and Mathematics Tests School Administrator’s Manual* (SAM), available online at <http://www.p12.nysed.gov/assessment/sam/ei/eisam17-v1.pdf> and <http://www.p12.nysed.gov/assessment/sam/ei/eisam17-v2.pdf>.

1.4. Test Use and Decisions Based on Assessment

The NYSTP Grades 3–8 ELA and Mathematics Tests are used to measure the extent to which individual students achieve the New York State CCLS in ELA and Mathematics, respectively, in order to determine whether or not schools, districts, and the State meet the required progress objectives specified in the New York State accountability system. Several types of scores are available from the Grades 3–8 ELA and Mathematics Tests, and they are discussed in this section.

1.4.1. Scale Scores

The scale scores are a quantification of the proficiency measured by the Grades 3–8 ELA and Mathematics Tests at each grade level. Scale scores are comparable only within a given subject and grade. Scale scores are not comparable across grades or across subjects. The scale scores are reported at the individual student level, and can be aggregated. Detailed information on the

derivation and properties of the scale scores is provided in Section 6: IRT Calibration and Linking. The Grades 3–8 ELA and Mathematics Tests’ scale scores are the basis for placing students into performance levels, which are used to determine student progress within schools and districts; support registration of schools and districts; determine eligibility of students for additional educational services; and provide teachers with indicators of a student’s need, or lack of need, for remediation in specific content-area knowledge.

1.4.2. *Statewide Percentile Ranks*

Students’ scale scores are also presented as percentile ranks in order to indicate student performance relative to the entire testing population on a scale that may be more familiar than the operational test’s scale. Such statistics are estimated based on how often each student earned a given scale score, thus presenting similar information as the scale score itself but on an alternate scale.

1.4.3. *Performance Level Cut Scores and Classification*

Student performance is classified as Level I, Level II, Level III, or Level IV for the Grades 3–8 ELA and Mathematics Tests. The definitions of performance levels are as follows:

- **NYS Level I:** Students performing at this level are well below proficient in standards for their grade. They demonstrate limited knowledge, skills, and practices embodied by the New York State P–12 Learning Standards for English Language Arts/Literacy or Mathematics that are considered insufficient for the expectations at this grade.
- **NYS Level II:** Students performing at this level are below proficient in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Learning Standards for English Language Arts/Literacy or Mathematics that are considered partial but insufficient for the expectations at this grade.
- **NYS Level III:** Students performing at this level are proficient in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Learning Standards for English Language Arts/Literacy or Mathematics that are considered sufficient for the expectations at this grade.
- **NYS Level IV:** Students performing at this level excel in standards for their grade. They demonstrate knowledge, skills, and practices embodied by the New York State P–12 Learning Standards for English Language Arts/Literacy or Mathematics that are considered more than sufficient for the expectations at this grade.

The performance level cut scores used to distinguish between Levels I, II, III, and IV were established during the process of standard setting in Summer 2013. The process is described in detail in Section 8 and Appendix P in the 2013 technical report (NYSED, 2013).

1.4.4. *Subscores*

The Grades 3–8 ELA tests have two subscores: reading (which includes all multiple-choice items assessing both reading and language standards) and writing to sources (which includes all constructed-response items assessing reading, writing, and language standards). The Grades 3–8

Mathematics tests have three subscores that are the domain-level scores for items measuring the *Major Clusters* in each grade. The CCLS are divided into *Major*, *Supporting*, and *Additional Clusters*. Standards within *Major Clusters* are the intended focus of instruction and assessment and account for the majority of the Mathematics test items. The *Supporting* and *Additional Clusters* are Mathematics standards that both introduce and reinforce *Major Clusters*. Tables 1.1 and 1.2 present the reporting subscore categories and the point values that correspond to each on the 2017 tests. In 2017, subscores were reported in two ways:

1. A raw score (i.e., number of points earned) out of the total score on the test
2. The average score at the state level for each subscore category

Table 1.1. ELA Subscore Categories and Total Possible Score Points

Grade	Total Subscore Points	
	Reading	Writing to Sources
3	25	22
4	25	22
5	35	22
6	35	22
7	35	22
8	35	22

Table 1.2. Mathematics Subscore Categories and Total Possible Score Points

Grade	Reporting Subscores and Total Subscore Points		
	Subscore 1	Subscore 2	Subscore 3
3	Operations and Algebraic Thinking 25	Number and Operations—Fractions 11	Measurement and Data 13
4	Operations and Algebraic Thinking 11	Numbers and Operations in Base 10 16	Number and Operations—Fractions 17
5	Numbers and Operations in Base 10 17	Number and Operations—Fractions 23	Measurement and Data 16
6	Ratios and Proportional Relationships 17	The Number System 15	Expressions and Equations 26
7	Ratios and Proportional Relationships 18	The Number System 15	Expressions and Equations 20

Grade	Reporting Subscores and Total Subscore Points		
	Subscore 1	Subscore 2	Subscore 3
8	Expressions and Equations 29	Functions 18	Geometry 16

1.5. Testing Accommodations

In accordance with federal law under the Americans with Disabilities Act and the section Fairness in Testing and Test Use in the *Standards for Educational and Psychological Testing* (AERA, APA, and NCME, 2014), accommodations that do not alter the measurement of any construct being tested are allowed for test takers. The allowance is in accordance with a student's Individualized Education Program (IEP) or Section 504 Accommodation Plan (504 Plan). School principals are responsible for ensuring that proper accommodations are provided when necessary, and that staff providing accommodations are properly trained. Details on testing accommodations can be found in the 2017 *NYSTP Grades 3–8 English Language Arts and Mathematics Tests School Administrator's Manual (SAM)*¹.

1.6. Test Transcriptions

For visually impaired students, large-type and Braille editions of the test books are provided. In most cases, the students dictate and/or record their responses, the teachers transcribe student responses to the multiple-choice items onto scannable answer sheets, and the teachers transcribe the responses to the constructed-response items onto the regular test books. Some of the students who use large-type editions will fill in the answer sheets by themselves. The large-type editions are created by Questar Assessment Inc. and printed by Midland Information Resources, and the Braille editions are produced by gh, LLC. gh employs certified Library of Congress Braille transcribers and delivers Braille in accordance with the Braille Authority of North America (BANA) standards. Camera-ready versions of the regular test books are provided to the Braille vendor, which then produces the Braille editions. Proofs of the Braille editions are submitted to NYSED for review and approval prior to production.

1.7. Test Translations

The NYSTP Grades 3–8 Mathematics Tests are translated into five languages: Chinese (Traditional), Haitian-Creole, Korean, Russian, and Spanish. These tests are translated to provide students the opportunity to demonstrate mathematical proficiency independent of their command of the English language. Sample tests are available in each translated language at the following location: <http://www.p12.nysed.gov/assessment/math/samplers/>.

English Language Learner/Multilingual Learner (ELL/MLL) students taking the Grades 3–8 Mathematics Tests may be provided with an oral translation of the test when a written translation is not available in the student's native language. The following testing accommodations are also made available to ELLs: separate testing location, bilingual glossaries, simultaneous use of English

¹ <http://www.p12.nysed.gov/assessment/sam/ei/eisam17-v1.pdf>
<http://www.p12.nysed.gov/assessment/sam/ei/eisam17-v2.pdf>

and alternative-language editions, oral translation for lower-incidence languages, and writing responses in the native language.

The NYSTP Grades 3–8 ELA Tests are not translated into any other language because they are assessments of proficiency in English language arts. The following testing accommodations are made available to ELLs taking the ELA Tests: separate testing location and bilingual glossaries.

Section 2: Test Design and Development

2.1. Test Descriptions

The 2017 Grades 3–8 ELA and Mathematics Tests are criterion-referenced tests composed of multiple-choice (MC) and constructed-response (CR) test items based on the New York State P–12 CCLS. The tests were administered in New York State classrooms during a three-day period from March to May of 2017. Details on the administration and scoring of these tests can be found in Section 4: Test Administration and Scoring. Additional information can be found in the *NYSTP Grades 3–8 English Language Arts and Mathematics Tests School Administrator’s Manual* (SAM).

2.1.1. ELA Tests

The 2017 Grade 3–8 ELA Tests were designed to measure student literacy as defined by the CCLS. The tests assessed Reading, Writing, and Language standards by using multiple-choice, short-response, and extended-response items. All items were based on close readings of informational, literary, or paired texts. All texts were drawn from authentic, grade-level works.

Multiple-choice items were designed to assess Reading and Language Standards. Multiple-choice items required students to analyze different aspects of a given text, including central idea, style elements, character and plot development, and vocabulary.

Short-response items were designed to assess Reading and Language Standards. These were single items in which students used textual evidence to support their answers to inferential questions. These items asked students to make an inference, state a position, or draw a conclusion based on their analysis of the passage and then provide two pieces of text-based evidence to support their answers. In responding to these items, students were expected to write in complete sentences. Appendix H provides the rubric for the short-response items.

Extended-response items were designed to assess Reading, Writing, and Language Standards, with a focus primarily on the Writing Standard. Extended-response items required comprehension and analysis of either an individual text or paired texts. Paired texts required students to read and analyze two related texts. Paired texts were related by theme, genre, tone, time period, or other characteristics. Many extended-response items asked students to express a position and support it with text-based evidence. For paired texts, students were expected to synthesize ideas between and draw evidence from both texts. Extended-response items required students to demonstrate their ability to write a coherent essay, using textual evidence to support their ideas. Appendix I provides the rubric for the extended-response items.

2.1.2. Mathematics Tests

The 2017 Grade 3–8 Mathematics Tests were designed to measure student mathematic understanding as defined by the CCLS. The tests required that students understand Mathematics conceptually, use prerequisite skills with grade-level mathematical facts, decide which formulas and tools (e.g., protractors and rulers) to use, and solve mathematics problems rooted in the real world. The tests contained multiple-choice, short-response (2-point), and extended-response (3-point) items. For multiple-choice items, students selected the correct response from four answer choices. For short- and extended-response items, students wrote an answer to an open-ended

question. Some items required students to show their work or to explain, in words, how they arrived at their answers.

Mathematics multiple-choice items were used mainly to assess standard algorithms and conceptual standards. Multiple-choice items incorporated the New York State CCLS, some in real-world applications. Many multiple-choice items required students to complete multiple steps. Likewise, many of these items were linked to more than one standard, drawing on the simultaneous application of multiple skills and concepts.

Short-response items were used mainly to assess conceptual and application standards. The items required students to complete a task and show their work. Like multiple-choice items, short-response items often required multiple steps and the application of multiple mathematics skills, some in real-world applications. Appendix J provides the rubric for the Mathematics short-response items.

Extended-response items were used mainly to assess students' abilities to show their understanding of mathematical procedures, conceptual understanding, and application of those procedures and concepts. Extended-response items required students to complete two or more tasks, or a more extensive problem, and show their work. Some items also assessed student reasoning and the ability to critique the arguments of others. Appendix K provides the rubric for the Mathematics extended-response items.

2.2. Test Configuration

2.2.1. Test Book Design

The 2017 Grades 3–8 ELA Tests were composed of three books per grade and administered in three sessions over three days. Each day consisted of one book: Book 1 and Book 2 contained literary and informational reading passages and MC items based on the passages. Book 2 also contained reading passages with short-response items and an extended-response item based on those passages. Book 3 contained only reading passages with short-response items and an extended-response item based on those passages.

The 2017 Grades 3–8 Mathematics Tests were composed of three books per grade and administered in three sessions over three days. Each day consisted of one book: Book 1 and Book 2 contained MC items. Book 3 contained short- and extended-response items. The tables in Appendix A provide information on the numbers and types of items in each book for the Grades 3–8 ELA and Mathematics Tests and the testing times.

2.2.2. Embedded Field-Test Items

In 2010, NYSED announced its commitment to embed multiple-choice items for field testing within the Spring 2012 Grades 3–8 ELA and Mathematics Operational Tests. This commitment continued for the Spring 2017 administrations of the tests. Embedding field-test items allows for a better representation of student responses and provides more reliable field-test data on which to build future operational tests. In other words, since the specific locations of the embedded field-test items were not disclosed and they look the same as operational test items, students were unable to differentiate field-test items from operational test items. Therefore, field-test data derived from embedded items are free of the effects of differential student motivation that may

characterize stand-alone field-test designs. Embedding field-test items also reduced the number of stand-alone field-test forms during Spring 2017, although it did not eliminate the need for them.

2.3. New York State Educators' Involvement in Test Development

New York State educators are actively involved in ELA and Mathematics test development. New York State educators provide critical input throughout all stages of the test development process, which include rangefinding, educator item review, operational forms construction, passage selection, item writing, and a “Final Eyes” meeting (a final review of the test books prior to printing).

In order to create fair and valid tests, NYSED gathers a diverse group of educators to review all test materials. The participants are selected for each testing activity based on:

- Certification and appropriate grade-level experience
- Special population experience
- Geographical region
- Gender
- Ethnicity
- Type of school (urban, suburban, or rural)

The selected participants must be certified and have both teaching and testing experience. Most of the participants are classroom teachers. Specialists such as reading coaches, literacy coaches, and special education and bilingual instructors also participate. Some participants are also recommended by principals, professional organizations, Big Four Cities (i.e., Buffalo, Rochester, Syracuse, and Yonkers), and/or the Staff and Curriculum Development Network (SCDN). A file of participants is maintained and routinely updated with current participant information, as well as the addition of possible future participants as recruitment forms are received. The process of continually updating and adding to this file contributes to NYSED's ability to include many educators in the test development process. Every effort is made to have diverse groups of educators participate in each testing event.

Additionally, Content Advisory Panels (CAPs) meet quarterly to review, vet, and provide comments on curricular and assessment work. CAPs are content-area-specific advisory panels composed of between 15 and 20 New York State P-12 educators whose members are nominated by state professional organizations, institutes of higher education, and educator unions.

2.4. Test Blueprints

After careful consideration of test length and administration constraints (e.g., location of multiple-choice and constructed-response items within test books), the representation and distribution of content were determined.

The CCLS for ELA are organized into four strands: Reading, Writing, Language, and Speaking/Listening. Due to administration constraints, Speaking/Listening was determined to best be assessed in the classroom, only; therefore, the ELA Tests assess three of the four strands: Reading, Writing, and Language. Content experts reviewed the Reading, Writing, and Language

standards and recommended content coverage by standard and item type, based on the depth and breadth of each standard.

The CCLS for Mathematics are divided into standards, clusters, and domains. Standards define what students should understand and be able to do and are further articulated into lettered components. Clusters are groups of related standards. Domains are larger groups of related clusters and standards. Content experts reviewed the Mathematics standards and recommended content coverage by standard and item type (MC or CR), based on the emphasis of the cluster (major, supporting, and additional) and depth and breadth of each standard.

Tables B1 and B2 in Appendix B show the test blueprint and actual number of score points in the Grades 3–8 ELA and Mathematics Tests, respectively. The tables include the ranges of allowable points for each ELA strand and Mathematics domain and the actual number of points on the 2017 operational tests.

2.5. Passage Selection and Item Criteria Documents

To guide test item development and to help ensure that New York State tests were measuring the CCLS for ELA and Mathematics with fidelity, criteria were established for selecting passages and writing test items, based on the consultation with the groups listed above.

The *Passage Selection Guidelines for Assessing State Standards (CCSS) ELA* were created to provide a framework that allows for the consistent selection of passages that are appropriately complex for the given grade and contain the specific characteristics necessary to measure different standards (see Appendix C). The guidelines describe the quantitative methods used to determine the grade appropriateness of a given text. They also describe the grade-specific text characteristics needed to develop items that measure any particular reading standard. The complete guidelines can be found here: http://www.engageny.org/sites/default/files/resource/attachments/passage_selection_guidelines_for_assessing_ccss_ela.pdf.

Passage Review Criteria documents were created based on the passage selection guidelines and were used to evaluate each potential passage and determine whether or not it could be used to measure the CCSS for ELA. The criteria documents were used to determine whether each passage suggested for testing use was grade appropriate, fair, and possessed the necessary characteristics to assess each standard. Specifically, passages were evaluated for the presence and quality of key ideas and details, craft and structure, and integration of knowledge and ideas. The full passage review criteria can be found here: http://www.engageny.org/sites/default/files/resource/attachments/new_york_state_passage_review_criteria_protocol_document.doc.

Item Review Criteria for the Grade 3–8 ELA Tests were used to help ensure that each item was clear and fair, measured a specific standard or standards with fidelity, and conformed to the specifications for each item type. Each section of the criteria includes pertinent questions used to determine whether or not an item was of sufficient quality so that it could move forward in the development process. The first two of the *Item Review Criteria*, clarity and fairness, identify the basic components of quality items. The criteria for clarity are used to help ensure that students understand what is asked in each item and that the language choice in the item does not negatively affect a student’s ability to perform the required task. For example, the criteria include checking to make sure that the vocabulary of test items is at grade level and that items avoid

technical terms unrelated to the content. Likewise, the fairness criteria are used to ensure that items are unbiased, non-offensive, and not disadvantageous to any given subgroup. The criteria also address how each item measures a given standard or standards and articulates the aspects of each standard that the items need to address. Finally, the criteria establish key requirements for each item type (e.g., requiring that each two-point constructed-response item asks students to make a clear statement that can be supported with two independent text-based pieces of evidence). The complete ELA criteria documents can be found here: <http://www.engageny.org/resource/new-york-state-item-review-criteria-for-grade-3-8-english-language-arts-tests>.

Item Review Criteria for the Grade 3–8 Mathematics Tests were used to ensure clarity, language and graphical appropriateness, fairness, freedom from bias, fidelity of measurement to the CCSS, and conformity to the expectations for specific item types and formats for each test item. Each section of the criteria includes pertinent questions that determine whether an item is of sufficient quality. The first two criteria, clarity and graphical appropriateness and fairness, identify the basic components of quality test items. The criteria for clarity and graphical appropriateness are used to help ensure that students understand what is asked in each item and that the language in the item does not adversely affect a student’s ability to perform the required task. For example, the criteria include checking to make sure that the visual load for any item containing art is reasonable and that interpreting a graphic does not confuse the underlying construct. Likewise, the fairness criteria are used to evaluate whether or not items are unbiased, non-offensive, and not disadvantageous to any given subgroup. The criteria also require documentation of how each item measures the assigned Mathematics standard(s). Finally, the criteria address the specific demands for different item types and formats (making sure that each three-point constructed-response item involves a multi-step process and requires students to show work). The complete Mathematics criteria document can be found here: <https://www.engageny.org/resource/new-york-state-item-review-criteria-for-grade-3-8-mathematics-tests>.

The *Multiple Representations for NYS Grade 3–8 Mathematics Tests* document was developed to ensure that the tests measured the deep conceptual understanding that the CCSS demands, rather than focusing on predictable Mathematics items that require only algorithmic strategies to be solved correctly. *Multiple Representations* is a broad set of specifications that describes, refers to, and symbolizes the various, but not all, ways that Mathematics standards could be measured within the constraints of the NYSTP. The document specifies three overarching families: procedural skills, conceptual understanding, and application. It also includes information about how to identify standards that might be measured through the use of a particular representation. It identifies types of Mathematics skills (e.g. application of process and explanation of a principle) that are appropriate for assessing different representations. The full document can be found here: <https://www.engageny.org/resource/multiple-representations-for-nys-grade-3-8-common-core-mathematics-tests>.

2.5.1. *Principles of Universal Design*

To create tests as equitable as possible for students, principles of Universal Design were employed during the creation of the tests and test items. In a report published by the National Council on Educational Outcomes, “‘Universally designed assessments’ are designed and developed from the beginning to allow participation of the widest possible range of students, and to result in valid inferences about performance for all students who participate in the assessment”

(Thompson, S.J., Johnstone, C.J., & Thurlow, M.L. 2002). The report goes on to describe seven elements of a universally designed assessment. These elements are:

1. Inclusive assessment population
2. Precisely defined constructs
3. Accessible, unbiased items
4. Amenable to accommodations
5. Simple, clear, and intuitive instructions and procedures
6. Maximum readability and comprehensibility
7. Maximum legibility

In accordance with these elements, the Universal Design Item Checklist in Appendix D was developed for use during item development.

2.6. Passage Finding

The goal of passage finding is to obtain high-quality texts from which to generate CCSS-aligned test items. To do so, in the 2015–2016 development cycle, independent passage finders were recruited and trained, using passage selection resources such as the passage selection criteria. Passage finders were given assignments based on the test blueprint requirements. Passage finders submitted passages along with completed criteria documents and source information to ELA content specialists, who reviewed the passages against the agreed-upon criteria. Passages that did not meet the criteria were rejected, and passages that did meet the criteria were moved forward in the process, where the text from scanned copies of the original sources was entered into templates. Once in the templates, readability metrics were determined for each text, and it was then proofread by copyeditors, fact checked by research librarians, reviewed for content issues by Science and Social Studies content specialists, and reviewed for Universal Design issues by specifically trained reviewers. After the passages went through these review steps, ELA content specialists posted the passages and completed criteria documents for NYSED’s review and approval for moving forward in the process.

NYSED staff retrieved and reviewed the passages and criteria documents. If NYSED staff determined that a passage did not meet the criteria, the passage was rejected and the NYSED staff provided an explanation for rejection.

In addition to the content reviews performed by NYSED staff and its vendors, the passages were also reviewed by executives in both organizations. The executive review focused on bias and sensitivity issues particular to New York State. Passages that passed both content and executive reviews were moved forward for item development.

2.7. Item Development

Item development for the 2017 test forms was conducted during the 2015–2016 development cycle. The goal of item development is to develop a sufficient number of high-quality, CCSS-aligned items to populate the test forms. Using the criteria documents for both content areas and the multiple-perspective document for Mathematics, content leads trained item writers. The item writers had teaching or assessment experience in the content area for which they were writing items; experience in writing for large-scale, high-stakes assessments; and, at minimum, a bachelor’s degree in either education and/or the content area for which they were assigned. The

item writers were given specific assignments, based on the test blueprint. For ELA, the item writers were also provided with the completed passage criteria documents.

Item writers provided items and completed criteria documents to content specialists for review. Two content specialists reviewed each item and its corresponding criteria document. Items that did not meet the criteria were sent back to the writers with specific feedback for revision. Items that did not meet the criteria after an attempted revision were rejected and content specialists replaced them. After the content specialists were satisfied that all of the items met the criteria, the items were reviewed by copyeditors. The Mathematics items were also reviewed by content specialists in Science and Social Studies and by research librarians. The ELA and Mathematics content specialists evaluated the feedback from the different internal groups and edited the items accordingly. The items and criteria documents were then posted for NYSED's review and approval for moving forward in the process.

NYSED content experts retrieved and reviewed the items and criteria documents. If NYSED staff determined that an item did not meet the criteria, the item was rejected and the NYSED staff provided an explanation for rejection. Questar content specialists then replaced the item and completed criteria documents, which were resubmitted to NYSED. If NYSED staff determined that an item met the criteria but could be improved with editing, the staff member recorded notes for the edits. Those notes were reviewed at face-to-face meetings at which content staff and NYSED staff reviewed and edited all of the items to ensure that they met the criteria. All passages and items accepted at that meeting were moved forward for the educator item review.

2.8. Educator Item Review

After being reviewed by NYSED, the items were presented to panels of New York State educators. Based on their expertise, educators were assigned to grade-level and content-specific groups where they reviewed the items. The reviews were facilitated by Questar content specialists and were attended by NYSED staff. For ELA, reviewers first read and then discussed the passages before reviewing items. For Mathematics and ELA, the educators used the following checklist to review each item.

1. Does the item align to the designated standard(s)?
 - The item measures the content standard(s) that it was designed to measure.
2. Does the item meet quality standards?
 - The item is worded clearly.
 - The reading level of the item is grade appropriate.
 - The item has one correct answer.
 - The item has plausible, unambiguous distractors.
 - All of the distractors are mutually exclusive.
3. Is the item fair?
 - The item is free from bias on the basis of students' personal characteristics, such as gender or ethnicity.

As the educators reviewed the items, they discussed their judgments about them. If the educators felt that an item did not align to the standards, did not meet quality standards, or was not fair, they made recommendations for editing the item. NYSED staff and Questar content specialists later reviewed the recommendations and made the appropriate edits.

2.9. Field-Testing

Once the items have been developed and thoroughly reviewed by a variety of stakeholders, they must then be field-tested. Field-testing items is a critically important step in the test development process, as it is only through the gathering of actual student response data that a variety of psychometric characteristics may be evaluated. Table 2.1 provides a summary of the unique items that passed the scrutiny of NYSED and Questar content specialists, as well as that of New York State educators, and were field-tested. More items were field-tested than were needed on the operational forms because that enabled tests to be constructed with items that include the best possible characteristics from both a content and psychometric perspective.

Table 2.1. Summary of Unique 2016 Field Test Items

Grade	Unique ELA Items by Type*		Unique Mathematics Items by Type*	
	MC	CR	MC	CR
3	110	35	90	34
4	111	34	90	34
5	106	33	90	32
6	112	35	100	34
7	106	35	100	31
8	112	35	100	28

* MC = multiple-choice. CR = constructed-response. All CR items were field-tested under stand-alone conditions, while MC items were administered under both embedded and stand-alone conditions.

Field-test items were administered in Spring 2016 as embedded field-test items within the 2016 operational test forms. The use of embedded field-test items yields more reliable field-test data and reduces, but does not eliminate, the need for multiple-choice stand-alone field-testing. One additional round of field-testing was administered separately from the 2016 operational forms (i.e., as stand-alone tests) later in Spring 2016.

In order to better understand how the field-test items may perform on future operational forms, a variety of analyses were conducted. All of the field-test data underwent a series of representativeness checks. Because only a small sample of schools participate for any given content area and grade for stand-alone field-testing, it was necessary to ensure that the stand-alone field-test samples were representative of the entire State population in terms of student achievement on prior years' tests, student gender, student ethnicity, and school Needs/Resource Capacity Category (NRC). Finally, a variety of psychometric analyses were conducted, including classical item analysis, inter-rater reliability for constructed-response items, differential item functioning (DIF), item response theory (IRT), item calibration, linking, scaling, and fit evaluation. Many of these analyses are described at length below. However, inter-rater reliability

analyses were not possible for the operational test, as only a single rater scored each constructed-response.

2.10. Rangefinding

Rangefinding for most items included on the 2017 test was conducted by Questar. Rangefinding occurs after constructed-response items have been field-tested. The purpose of rangefinding is to have New York State educators review student constructed-responses and arrive at consensus scores based on the standards established by NYSED and the scoring rubrics. The consensus scores become the basis for operational rating guides and scoring ancillaries. To arrive at consensus, committees of New York State educators review, discuss, and rate student responses to the constructed-response field-test items. This process was overseen by NYSED content experts and Questar Scoring Directors. The first step in the rangefinding process was to have the educator committees review rubrics and a NYSED-approved grounding guide set, previously used for the 2016 field-test rangefinding sessions, to familiarize teachers with the application of NYSED standards and rubrics. The grounding guide sets contain student responses that illustrate the full range of scores on the rubric. The grounding guide sets are composed of student responses that had previously gone through the rangefinding process and been approved by NYSED, and are used to guide the scoring of field-test and operational student responses. Referencing the previously approved guide set papers during the rangefinding sessions ensures consistency in the application of NYSED standards and rubrics from year-to-year.

After the committee reviewed the pre-approved grounding guide set, groups of committee members familiarized themselves with each item type, scoring a small number of responses representative of each of the different score points. After the group-scoring exercise, committee members independently scored other student responses. The committee then reviewed and discussed their results and determined consensus scores for the responses. The rangefinding results were used to build training materials for Questar scorers, who scored the field-test responses to constructed-response items.

2.11. Item Selection and Test Creation (Criteria and Process)

The NYSTP Grades 3–8 ELA and Mathematics Tests were administered from March to May of 2017. The test items were selected from the pools of available ELA and Mathematics items. These items were field-tested either in embedded field-testing or stand-alone field-testing from 2013 through 2015.

The test construction process involved several iterative steps. Three criteria governed the item selection process:

- Meet the ELA and Mathematics content specifications provided by NYSED
- Select items with the best psychometric characteristics from the ELA and Mathematics item pools
- Combine psychometric characteristics of all selected items with the intended psychometric goals for each entire form

Questar content specialists were provided the test designs, blueprints, and psychometric guidelines for item selection. The psychometric guidelines were based on the classical and IRT statistics associated with the test items.

Using the pool of field-tested items, Questar content specialists made preliminary selections for each grade and content area. The selections were then reviewed by the content leads for each content area, to make sure that the items conformed to the different criteria. If the content criteria were not met, new items were selected. After the content leads' review, the item selections were reviewed by Questar psychometricians. If items with undesirable statistics were selected, the psychometricians proposed items with more desirable statistics. Those items were then reviewed by the content specialists and their leads. Once the Questar content teams and the psychometric teams were satisfied that the content and statistics of the selected items and the proposed whole forms met the requirements, the items were given to NYSED staff (including content and assessment experts) to review. Questar content specialists and psychometricians traveled to Albany, New York, in October 2016 to finalize item selection and test creation with NYSED staff (including content and assessment experts) and educators.

2.12. Educator Form Construction

During an educator form construction meeting that took place from November 7–November 18, 2016 (excluding November 10, as it was a holiday), in Albany, New York, educators from around the State worked with NYSED and Questar to review the content of the proposed 2017 operational ELA passages, and ELA and Mathematics individual test items, and how those items combine to create entire operational forms, for quality and appropriateness using their subject matter expertise. The goal was to ensure that all test items and forms are defensible from content and psychometric perspectives. The outcome was test forms that meet psychometric parameters and contain items that meet content criteria.

A different group of educators participated in the review of each subject and grade's test form, so each morning began with training in each room. Once training was complete, participants began the form construction process by independently evaluating the items and passages (for ELA) against the criteria on the provided checklists. Each participant completed his or her own checklist and had a binder with item cards corresponding to the order of items in the test.

- For ELA, the educators initially reviewed the first passage and a single item from the passage. Once they got used to the process, the educators reviewed the passages and the corresponding items. During this review, educators confirmed that there was only one correct answer for each multiple-choice item, and that the item was aligned to the standard that it purported to address. They also estimated the time that it would take for students to read the passage and answer the items.
- For Mathematics, the educators initially reviewed single items and discussed each item as a group. Once they got used to the process, the educators reviewed groups of items (e.g., 4 to 6 items, followed by discussion of each item). During this review, educators confirmed that there was only one correct answer for each multiple-choice item, and that the item was aligned to the standard that it purported to address. They also estimated the time that it would take for students to answer the items.

In both ELA and Mathematics, the educators, in consultation with NYSED and Questar content experts, were permitted to recommend:

- revisions to the stated standard alignment;
- revisions to item sequencing to avoid cueing/clueing; and
- swapping any items and/or passages that they judged as having problems flagged by the above reviews.

Given other constraints, it was not always possible to make every change that educators recommended, but they were given the opportunity to voice any and all concerns that they had and NYSED made the final decision about any educator recommendations.

The facilitators then led a group discussion and helped the group reach consensus. Where time permitted, educators were presented with and approved the items that Questar and NYSED proposed for any necessary replacements. Following each session with educators, NYSED and Questar met to review the content and data of the proposed selections, and explore alternate selections for consideration. NYSED then approved the item selections, including item positions within test books.

2.13. Test Form Production

Once the selection of items for the operational and embedded field-test positions was completed, Questar created test forms. The test forms were reviewed by Questar content specialists and were posted for NYSED to review. NYSED and Questar reviewed the forms to look for any errors in spelling, capitalization, punctuation, grammar, and formatting. They also confirmed that each multiple-choice item had a single correct answer.

2.14. Final Eyes Committees

After NYSED and Questar reviewed copies of the test forms, the test forms were reviewed by the Final Eyes committees. For each content area, the committee consisted of nine New York State educators from around the State. During that review, the educators were charged with taking the test to make sure that each multiple-choice item had a single correct answer, and to look for errors in spelling, capitalization, punctuation, grammar, and formatting.

After the Final Eyes review and after NYSED approved edits made as a result of the review, the tests were then considered final and produced for the 2017 administration.

2.15. Proficiency and Performance Standards

In Summer 2013, after the operational administration of the 2013 tests, a standard setting meeting occurred in Albany where 95 New York State educators went through a rigorous process, guided by the best practices indicated by this intensely studied process, to recommend performance standards for the new tests measuring the CCLS. These recommendations were presented to the Commissioner and the Board of Regents, who, in turn, adopted the recommended standards set forth by the committees. For additional details, see Section 8 and Appendix P in the 2013 technical report (NYSED, 2013).

Each grade level has four performance levels. Three cut points demarcate the performance levels needed to demonstrate each ascending level of performance. Section 6.8.1. Raw Score-to-Scale

Score and SEM Conversion Tables contains detailed information related to performance standards.

Section 3: Validity

Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses of tests. Test validation is an ongoing process of gathering evidence from many sources to evaluate the soundness of the desired score interpretation or use. This evidence is acquired from studies of the content of the test and studies involving scores produced by the test. Additionally, reliability has to be considered before considerations of validity are made. A test cannot be valid if the test scores are not first reliable.

The *Standards for Educational and Psychological Testing* (AERA, APA, and NCME, 2014) addressed the concept of validity in testing, which refers to the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores. Validity is the most important consideration in test evaluation. Test validation is the process of accumulating evidence to support any particular inference. Validity, however, is a unitary concept. Although evidence may be accumulated in many ways, validity refers to the degree to which evidence supports the inferences made from test scores.

3.1. Content Validity

Generally, achievement tests are used for student-level outcomes, either for making predictions about students or for describing students' performances (Mehrens and Lehmann, 1991). Tests are now also used for the purposes of accountability and adequate yearly progress (AYP). The NYSED uses various assessment data in reporting AYP. Specific to student-level outcomes, the NYSTP documents student performance in the area of Mathematics as defined by the New York State Mathematics Learning Standards and in the area of ELA as defined by the New York State ELA Learning Standards.

To allow test score interpretations appropriate for this purpose, the content of the test must be carefully matched to the specified standards. The 2014 AERA/APA/NCME standards state that content-related evidence of validity is a central concern during test development. Expert professional judgment should play an integral part in developing the definition of what is to be measured, such as describing the universe of the content, generating or selecting the content sample, and specifying the item format and scoring system.

Expert analysis of test content indicates the degree to which the content of a test covers the domain of content that the test is intended to measure. In the case of the NYSTP, the content is defined by detailed blueprints that describe New York State content standards and define the skills that must be measured to assess these content standards (see Tables B1 and B2 in Appendix B). The NYSTP test development process requires specific attention to content representation and the balance within each test form. New York State educators were involved in test construction in various development stages. For example, during the item review process, they reviewed field-test items for the alignment of the items with the CCLS. Educators also participated in a process of establishing scoring rubrics for constructed-response items during rangefinding. Section 2: Test Design and Development contains more information specific to the item review process.

3.2. Construct (Internal Structure) Validity

Construct validity (i.e., what scores mean and what kind of inferences they support) is often considered the most important type of test validity. Construct validity of the NYSTP Grades 3–8 ELA and Mathematics Tests are supported by several types of evidence that can be obtained from the ELA and Mathematics test data.

3.2.1. Internal Consistency

Empirical studies of the internal structure of the test provide one type of evidence of construct validity. For example, high internal consistency constitutes evidence of validity. This is because high coefficients imply that the test items are measuring the same domain of skill and are reliable and consistent. Reliability coefficients of the tests for total populations and subgroups of students are presented in Section 7.1: Test Reliability. For the total population, the ELA reliability coefficients (Cronbach’s alpha) ranged from .89 to .92. For all subgroups, the reliability coefficients were greater than or equal to .78. For the total population, the Mathematics reliability coefficients (Cronbach’s alpha) ranged from .93 to .95. For all subgroups, the reliability coefficients were greater than or equal to .79. Overall, high internal consistency of the NYSTP Grades 3–8 ELA and Mathematics Tests provided sound evidence of construct validity.

3.2.2. Unidimensionality

Other validity evidence comes from analyses of the degree to which the test items conform to the requirements of the statistical models. These statistical models are used to scale and link the tests, as well as to generate student scores. The models require that the items fit the model well (item fit) and that the items in a test measure a single domain of skill (unidimensionality).

The first step is to assess the degree to which the items fit the IRT model. The item-model fit for the ELA and Mathematics tests was assessed using Q_I statistics (Yen, 1981), and the results are described in detail in Section 6: IRT Calibration and Linking. Most items demonstrated sound fit across grades and content areas, and only a few items were deemed to have less than ideal fit. This provides solid evidence for the appropriateness of the IRT models used to calibrate and scale the test data.

Additional evidence for the efficacy of the model involves demonstrating that the items on the New York State tests are related to each other, within their respective content areas. This relationship of the items within the ELA or Mathematics tests is the common proficiency acquired by students studying the content area. This “common proficiency,” or, more formally, underlying construct, could be labeled as ELA proficiency (using the ELA scores) or Mathematics proficiency (using the mathematics scores), depending on the degree to which the ELA and Mathematics items are related.

Factor analysis of the test data is one way of modeling the common construct. This analysis may show that there is a single or main factor that can account for much of the variability between responses to test items. A large first component in factor analysis would provide evidence of the latent proficiency that students have in common regarding the particular items asked. A large main factor found from a factor analysis of an achievement test would suggest a primary construct that may be related to what the items were designed to have in common (i.e., Mathematics proficiency or ELA proficiency).

To demonstrate the common factor underlying student responses to the ELA and Mathematics test items, principal component factor analyses were conducted on a correlation matrix of individual items for the ELA and Mathematics tests. Factoring a correlation (i.e., tetrachoric correlation) matrix rather than actual item response data is preferable when dichotomous variables are in the analyzed data set. Because the ELA and Mathematics tests contain both multiple-choice and constructed-response items, the matrices of *polychoric* correlations were used as input for the factor analyses, as polychoric correlations are appropriate with both multiple-choice and constructed-response data. The study was conducted on the New York State public, charter, and religious and independent school students for whom data were available during the linking process. A large first principal component was evident in each analysis, demonstrating essential unidimensionality of the trait (i.e., proficiency) measured by each test. In other words, statistical evidence indicates that the ELA items are measuring one underlying construct, ELA proficiency, and that the Mathematics items are measuring one underlying construct, Mathematics proficiency.

The factor analyses conducted with the ELA and Mathematics data will show almost as many underlying constructs, or factors, as there are items on the test. Therefore, it is necessary to further investigate the factor analysis results to determine the number of “meaningful” factors. Specifically, more than one factor with an eigenvalue greater than 1.0 present in each dataset would suggest the presence of small additional factors. The magnitude of the ratio of the variance accounted for by the first factor compared to the remaining factors also provides evidence as to the number of meaningful factors. In addition, the total amount of variance accounted for by the main factor was evaluated. According to M. Reckase (1979),

“... the 1PL and the 3PL models estimate different abilities when a test measures independent factors, but ... both estimate the first principal component when it is large relative to the other factors. In this latter case, good ability estimates can be obtained from the models, even when the first factor accounts for less than 10 percent of the test variance, although item calibration results will be unstable.”

Factor analyses related to the Grades 3–8 ELA and Mathematics Tests indicated that the ratio of the variance accounted for by the first factor to the remaining factors was sufficiently large to support the claim that the ELA and Mathematics tests were essentially unidimensional; the ELA-related ratios and the Mathematics-related ratios showed that the first eigenvalues were at least five times as large as the second eigenvalues for all of the grades.

All of the Grades 3–8 ELA and Mathematics Tests exhibited first principal component accounting for more than 19% and 29% of the test variance, respectively. Tables 3.1 and 3.2 present the results of factor analyses, including eigenvalues greater than 1.0 and proportions of variance explained by the extracted factors, for ELA and Mathematics, respectively.

The evidence in Table 3.1 supports the claim that one single construct underlies the items/tasks in each ELA test and that scores from each test would represent performance primarily determined by that construct. Construct-irrelevant variance does not appear to create significant nuisance factors. Similarly, Table 3.2 supports the claim that a common construct underlies the items/tasks in each Mathematics test and that scores from each test would represent performance

primarily determined by that construct. Construct-irrelevant variance does not appear to create significant nuisance factors.

Table 3.1. ELA Tests Factor Analysis

Grade	Extracted Factor			
	#	Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
3	1	8.68	25.52	25.52
	2	1.50	4.40	29.92
	3	1.02	3.00	32.92
4	1	7.56	22.22	22.22
	2	1.42	4.19	26.41
	3	1.08	3.16	29.57
	4	1.01	2.98	32.55
5	1	8.39	19.06	19.06
	2	1.47	3.33	22.40
	3	1.15	2.61	25.00
	4	1.05	2.38	27.39
	5	1.01	2.31	29.69
6	1	10.10	22.96	22.96
	2	1.92	4.36	27.32
	3	1.17	2.67	29.99
	4	1.03	2.33	32.32
7	1	9.52	21.64	21.64
	2	2.01	4.57	26.21
	3	1.22	2.77	28.98
	4	1.03	2.34	31.32
8	1	9.77	22.19	22.19
	2	1.86	4.23	26.42
	3	1.25	2.83	29.26
	4	1.04	2.36	31.62

Table 3.2. Mathematics Tests Factor Analysis

Grade	Extracted Factor			
	#	Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
3	1	12.38	27.51	27.51
	2	1.76	3.91	31.42
	3	1.13	2.51	33.93
	4	1.05	2.33	36.26
	5	1.03	2.29	38.55
4	1	13.57	28.28	28.28
	2	1.76	3.66	31.93
	3	1.14	2.38	34.32
	4	1.09	2.26	36.58
	5	1.01	2.09	38.67
5	1	13.90	28.96	28.96
	2	1.77	3.68	32.65
	3	1.13	2.35	34.99
6	1	14.12	26.14	26.14
	2	1.72	3.18	29.32
	3	1.17	2.16	31.48
7	1	15.62	28.92	28.92
	2	1.60	2.96	31.88
	3	1.09	2.02	33.90
8	1	12.73	23.58	23.58
	2	1.58	2.92	26.50
	3	1.21	2.24	28.74
	4	1.05	1.95	30.69

As additional evidence for construct validity, the same factor analysis procedure was employed to assess the dimensionality of the Mathematics construct for selected subgroups of students in each grade: English language learners (ELLs), students with disabilities (SWD), and students using test accommodations (SUA), as well as ELL/SUA, and SWD/SUA. The ELL/SUA subgroup is defined as examinees who are ELLs and who use at least one ELL-related accommodation. The SWD/SUA subgroup includes examinees who are classified as having disabilities and who use at least one disability-related accommodation. The results were comparable to the results obtained from the total population data. Evaluation of eigenvalue magnitude and proportions of variance explained by the main and secondary factors provide evidence of essential unidimensionality of the construct measured by the tests for the analyzed subgroups. Appendix L provides factor analysis results for ELL, SWD, SUA, ELL/SUA, and SWD/SUA classifications.

3.2.3. *Detection of Bias*

Minimizing item bias has the goal of minimizing construct-irrelevant variance and helps establish a strong validity argument for the tests. Specifically, bias occurs if items function differentially for key pairs of groups, which may, in turn, cause the test to be differentially valid for certain groups of test takers. The statistical means for flagging items that may exhibit bias is referred to as differential item functioning (DIF). These statistical procedures were designed to be conservative (i.e., they were designed to flag more items for DIF, rather than fewer). Therefore, it is rare in practice to observe a high-stakes test in which not a single item is flagged for DIF. Since these procedures tend to over-flag items, it is only through review of those flagged items by experts that the items flagged for DIF may be judged to have or be free of bias. If the test involves irrelevant skills or knowledge, the possibility of bias is increased. Thus, preserving content validity is essential.

The developers of the NYSTP tests gave careful attention to items of possible ethnic, gender, socioeconomic status (SES), and—only for the Mathematics tests—translation bias. All materials were written and reviewed to conform to Questar’s editorial policies and guidelines for equitable assessment, as well as NYSED’s guidelines for item development. All materials were written to NYSED’s specifications and carefully checked by groups of trained New York State educators during the item review process. These steps are essential in keeping bias to a minimum. However, current evidence suggests that expertise in this area is no substitute for data; reviewers are sometimes wrong about which items work to the disadvantage of a group, apparently because some of their ideas about how students will react to items may be faulty (Sandoval & Mille, 1979; Jensen, 1980). Thus, empirical studies were conducted.

Statistical methods were used to identify items exhibiting possible DIF. Although items flagged for DIF in the field-test stage were closely examined for content bias and avoided during the operational test construction, DIF analyses were conducted again on operational test data. Different methods were employed to evaluate the amount of DIF in all test items: constructed-response items were evaluated with standardized mean differences, and multiple-choice items were analyzed using Mantel-Haenszel methods (see Section 5: Operational Test Data Collection and Classical Analysis).

In each grade, for both ELA and Mathematics, few items were flagged for DIF. Moreover, the magnitude of DIF for the flagged items was typically small (for more details, see Appendix N). Items flagged for statistically significant DIF were carefully reviewed by multiple reviewers during the operational test item selection. All such items were deemed by the reviewers to be free of bias (i.e., judged not to adversely affect any demographic subgroup studied) and remained in the tests.

Section 4: Test Administration and Scoring

This section provides summaries of New York State test administration and scoring procedures. For further information, refer to the aforementioned *School Administrator's Manual* and the *New York State Scoring Leader Handbook (2017)* located here: <http://www.p12.nysed.gov/assessment/sam/ei/scoringleaderhb17.pdf>.

4.1. Test Administration

The NYSTP Grades 3–8 ELA and Mathematics Tests were administered to students in a paper-based (PBT) and computer-based (CBT) testing mode in 2017. The PBT window was Tuesday, March 28–Thursday, March 30 for the Grades 3–8 ELA Tests and Tuesday, May 2–Thursday, May 4 for the Grades 3–8 Mathematics Tests. The CBT window was Monday, March 27–Monday, April 3 for the Grades 3–8 ELA Tests and Monday, May 1–Monday, May 8 for the Grades 3–8 Mathematics Tests.

The makeup test administration windows allowed students who were ill or otherwise unable to test during the assigned window to take the tests. The makeup test administration window for the paper-based test was Friday, March 31–Wednesday, April 5 for the Grades 3–8 ELA Tests and Friday, May 5–Wednesday, May 10 for the Grades 3–8 Mathematics Tests. The makeup test administration window for the Computer-based test was Tuesday, April 4–Thursday, April 6 for the Grades 3–8 ELA Tests and Tuesday, May 9–Thursday, May 11 for the Grades 3–8 Mathematics Tests.

4.2. Scoring Procedures of Operational Tests

The scoring of the NYSTP 2017 Grades 3–8 ELA and Mathematics Tests was performed at designated sites by qualified teachers and administrators. The number of personnel at a given site varied, as districts have the option of regional, district-wide, or school-wide scoring (please refer to Section 4.3: Scoring Models for more details). Administrators were responsible for the oversight of scoring operations, including the preparation of the test site, the security of test books, and the supervision of the scoring process. At each site, designated trainers taught scoring committee members the basic criteria for scoring each item and monitored the scoring sessions in the room. The trainers were assisted by facilitators or leaders, who also helped in monitoring the sessions and enforced scoring accuracy.

The titles for administrators, trainers, and facilitators vary by the scoring model that is selected. At the regional level, oversight was conducted by a site coordinator. A scoring leader trained the scoring committee members and monitored the sessions, and a table facilitator assisted in monitoring the sessions. For each subject, the oversight was structured in the same way for district- and school-wide models. At the district-wide level, a school district administrator oversaw scoring. A district subject leader trained the scoring committee members and monitored the sessions, and a school subject leader assisted in monitoring the sessions. For school-wide scoring, oversight was provided by the principal; otherwise, titles for the school-wide model were the same as those for the district-wide model. The general title “scoring-committee members” included scorers at every site.

4.3. Scoring Models

For the 2016–2017 school year, schools and school districts were able to score Grades 3–8 ELA and/or Mathematics Tests regionally, multi-district, district-wide, or school-wide, based on local need. Schools were required to enter one of the following scoring model codes on student answer sheets:

1. Regional scoring—The scorers for the school’s test papers included either staff from three or more school districts or staff from all religious and independent schools in an affiliation group (religious and independent or charter schools may participate in regional scoring with public school districts, and may be counted as one district).
2. Schools from two districts—The scorers for the school’s test papers included staff from two school districts, religious and independent schools, charter school districts, or a combination thereof.
3. Three or more schools within a district—The scorers for the school’s test papers included staff from all schools administering this test in a district, provided that at least three schools are represented.
4. Two schools within a district—The scorers for the school’s test papers included staff from all schools administering this test in a district, provided that two schools are represented (not available for CBT schools).
5. One school, only (local scoring)—The first readers for the school’s test papers included staff from the only school in the district administering this test, staff from one charter school, or staff from one religious and independent school (not available for CBT schools).
6. Private contractor—Scored by a private contractor that does not belong to Boards of Cooperative Educational Services (BOCES).

Schools and districts were instructed to carefully analyze their individual needs and capacities to determine their appropriate scoring model. BOCES and the Staff and Curriculum Development Network (SCDN) provided districts with technical support and advice in making this decision.

4.4. Scoring of Constructed-Response Items

The key resource for both the training of scoring committee members and the scoring of CR items was the scoring guides. These documents were created by Questar from sets of actual field-test student responses that were consensus scored by NYSED and New York State teachers during Rangefinding sessions. Trainers used these materials to train scoring-committee members on the criteria for scoring CR items. Additionally, scoring leader handbooks were also distributed to outline the responsibilities of the scoring roles. Both CBT and PBT responses were hand-scored by this process.

Upon completion of the training of scoring committee members, scoring was conducted with pen-and-pencil scoring as opposed to electronic scoring, and each scoring-committee member evaluated actual student papers instead of electronically scanned papers for PBT responses. CBT responses were evaluated electronically. All scoring-committee members were trained by previously trained and approved trainers along with guidance from scoring guides. Each constructed-response test book was scored by three separate scoring committee members, who scored three distinct sections of the test book. In order to verify the accuracy of scoring, after test

books were completed, the table facilitator or subject (ELA or mathematics) leader conducted a “read behind” of approximately 12 sets of test books per hour. If an item arose that was not covered in the training materials, facilitators or trainers were to call the Questar Scoring Helpline for assistance with the ELA or mathematics scoring (see Section 4.6. Quality Control Process).

4.5. Scorer Qualifications and Training

The scoring of the 2017 Grades 3–8 ELA and Mathematics Tests was conducted by qualified administrators and teachers. Trainers used the scoring guides to train scoring-committee members on the criteria for scoring constructed-response items. Part of the training process was the administration of a consistency assurance set (CAS) that provided the State’s scoring sites with information regarding strengths and weaknesses of their scorers. This tool allowed trainers to retrain their scorers, if necessary. The CAS also acknowledged those scorers who had grasped all aspects of the content area being scored and were well prepared to score student responses.

Regardless of the scoring model used, a minimum of three scorers is necessary to score each student’s test. However, to comply with a State requirement, none of the scorers assigned to score a student’s test responses may be that student’s teacher. This policy is detailed in the *Scoring Leader Handbook* section “Assigning Scorer Numbers and Questions to Scoring Committee Members” on page 21, found online at: <http://www.p12.nysed.gov/assessment/sam/ei/scoringleaderhb17.pdf>.

4.6. Quality Control Process

Test books were randomly distributed throughout each scoring room so that books from each region, district, school, or class were evenly dispersed. Teams were divided into groups of three, in order to ensure that a variety of scorers graded each book. If a scorer and a facilitator could not reach a decision on a paper after reviewing the scoring guides and audio files, they called the Questar Scoring Helpline. The call center was established to help teachers and administrators during scoring. The helpline staff consisted of trained Questar personnel, who answered questions by phone or fax. When a member of the staff was unable to resolve an issue, it was referred to NYSED for a scoring decision. In order to certify that all of the items were scored and that the scoring-committee members darkened each score on the answer document appropriately, a quality check was also performed on each completed box of scored tests. The log of calls received by the scoring helpline was delivered to NYSED twice daily during the scoring window. To affirm that all schools across the state adhered to scoring guidelines and policies, approximately 5% of the schools’ results are audited each year by an outside vendor.

Section 5: Operational Test Data Collection and Classical Analysis

5.1. Data Collection

Test data were collected in two phases. During Phase 1, a sample of approximately 95% of the student test records were received from the data warehouse and delivered to Questar, beginning at the end of May 2017. During Phase 2, “straggler files” were submitted to Questar in June 2017.

The “straggler files” contained fewer than about 5% of the total population cases, and were excluded from the classical, IRT, and reliability analyses (as described in Sections 5, 6, and 7, respectively) due to late submission. The analyses described in Section 8, “Summary of Operational Test Results,” were based on the data collected from both Phase 1 and Phase 2. Data collected from both public schools and religious and independent schools were included in all data analyses.

5.2. Data Processing

Depending on the nature of the analysis, more student records were included in some analyses than in others. For example, all students with valid test scores were included in the analyses described in Section 8, “Summary of Operational Test Results.” For the analyses described in other sections, however, more stringent data cleaning procedures were applied (see details below).

Data processing here refers to the cleaning and screening procedures used to identify errors (such as out-of-range data), and the decisions made to exclude student cases or to suppress particular items in certain analyses. In order to obtain a sample of the utmost integrity, Questar’s psychometric team performed data cleaning to the delivered data, and excluded some student cases. It should be noted that a student case being excluded from certain data analyses did not mean that the student record was invalidated. According to the NYSED’s specific instructions, additional procedures were taken to correct or recover these students’ records so that their test results were scored properly. As mentioned above, their records were included in later analyses (see Section 8).

The major groups of cases excluded from the data set (used for analyses in Sections 5, 6, and 7) were students with missing school type and those with at least one entirely missing test book. Other deleted cases included students with incorrect or incomplete grade information, duplicate record cases, and no-response record cases. The mathematical data cleaning procedure also excluded records with mismatched form language indicators for translated versions across the three test books for a given student.

5.2.1. Sampling Down for Representativeness

Historically, after data cleaning, the sample is reviewed for representativeness of the prior year’s operational population (i.e., all students testing in Spring 2016) in terms of key variables such as student gender, racial/ethnic identity, student disability status, English Language Learner (ELL) status, presence of test accommodation(s), and school Needs/Resource Capacity Category (NRC). At the recommendation of New York State’s Assessment Technical Advisory Committee (TAC), Questar shifted the focus from sampling down according to demographic representativeness, to instead focus on matching the prior year’s population’s distribution of ability. Questar and NYSED still reviewed the demographic patterns for 2017 relative to 2016,

but they were not used directly in the sampling down analyses. Comparison results between the final 2017 sample and 2016 operational population are further described in Section 6: IRT Calibration and Linking. In Spring 2017, a sampling down approach was adopted to make the sample used for linking as similar as possible to the previous year's testing population.

The numbers of cases considered for dropping because of sampling down varied across grades and subjects, but the process for all grades was consistent. The cleaned data file for a given subject and grade was the starting point. Questar reviewed the distribution of raw score proportion correct (RSPC) for the 2016 and 2017 operational forms. There were some minor differences in the 2016 and 2017 distributions of RSPC, but overall Questar, NYSED, and its TAC agreed that there was no evidence for a need to sample down in any subject or grade.

The data cleaning procedures and accompanying case counts are represented for ELA and Mathematics in Tables 5.1–5.6 and Tables 5.7–5.12, respectively.

Table 5.1. ELA Grade 3 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	206,560
Wrong Subject	0	206,560
No Grade	1	206,559
Wrong Grade	117	206,442
Language or Mismatched Form	3,908	202,534
School Type	125	202,409
Missing Entire Book	27,264	175,145
Invalid Score	15	175,130
Out-of-Range CR Scores	0	175,130
Duplicated Record	32	175,098

Table 5.2. ELA Grade 4 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	211,846
Wrong Subject	0	211,846
No Grade	0	211,846
Wrong Grade	107	211,739
Language or Mismatched Form	4,231	207,508
School Type	140	207,368
Missing Entire Book	32,519	174,849
Invalid Score	16	174,833
Out-of-Range CR Scores	0	174,833
Duplicated Record	12	174,821

Table 5.3. ELA Grade 5 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	205,052
Wrong Subject	0	205,052
No Grade	1	205,051
Wrong Grade	123	204,928
Language or Mismatched Form	4,066	200,862
School Type	152	200,710
Missing Entire Book	36,098	164,612
Invalid Score	10	164,602
Out-of-Range CR Scores	0	164,602
Duplicated Record	6	164,596

Table 5.4. ELA Grade 6 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	204,715
Wrong Subject	0	204,715
No Grade	0	204,715
Wrong Grade	138	204,577
Language or Mismatched Form	2,827	201,750
School Type	282	201,468
Missing Entire Book	40,028	161,440
Invalid Score	2	161,438
Out-of-Range CR Scores	0	161,438
Duplicated Record	14	161,424

Table 5.5. ELA Grade 7 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	201,266
Wrong Subject	0	201,266
No Grade	2	201,264
Wrong Grade	160	201,104
Language or Mismatched Form	2,672	198,432
School Type	225	198,207
Missing Entire Book	45,854	152,353
Invalid Score	7	152,346
Out-of-Range CR Scores	0	152,346
Duplicated Record	8	152,338

Table 5.6. ELA Grade 8 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	203,934
Wrong Subject	0	203,934
No Grade	0	203,934
Wrong Grade	111	203,823
Language or Mismatched Form	2,340	201,483
School Type	186	201,297
Missing Entire Book	58,076	143,221
Invalid Score	8	143,213
Out-of-Range CR Scores	0	143,213
Duplicated Record	6	143,207

Table 5.7. Mathematics Grade 3 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	209,652
Wrong Subject	0	209,652
No Grade	0	209,652
Wrong Grade	37	209,615
Language or Mismatched Form	2,735	206,880
School Type	142	206,738
Missing Entire Book	28,593	178,145
Invalid Score	21	178,124
Out-of-Range CR Scores	0	178,124
Duplicated Record	38	178,086

Table 5.8. Mathematics Grade 4 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	213,759
Wrong Subject	0	213,759
No Grade	0	213,759
Wrong Grade	27	213,732
Language or Mismatched Form	3,154	210,578
School Type	162	210,416
Missing Entire Book	33,721	176,695
Invalid Score	4	176,691
Out-of-Range CR Scores	0	176,691
Duplicated Record	12	176,679

Table 5.9. Mathematics Grade 5 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	208,044
Wrong Subject	0	208,044
No Grade	1	208,043
Wrong Grade	35	208,008
Language or Mismatched Form	2,877	205,131
School Type	181	204,950
Missing Entire Book	38,367	166,583
Invalid Score	3	166,580
Out-of-Range CR Scores	0	166,580
Duplicated Record	10	166,570

Table 5.10. Mathematics Grade 6 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	208,495
Wrong Subject	0	208,495
No Grade	2	208,493
Wrong Grade	40	208,453
Language or Mismatched Form	3,425	205,028
School Type	142	204,886
Missing Entire Book	42,892	161,994
Invalid Score	6	161,988
Out-of-Range CR Scores	0	161,988
Duplicated Record	18	161,970

Table 5.11. Mathematics Grade 7 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	196,628
Wrong Subject	0	196,628
No Grade	0	196,628
Wrong Grade	51	196,577
Language or Mismatched Form	2,633	193,944
School Type	213	193,731
Missing Entire Book	50,806	142,925
Invalid Score	9	142,916
Out-of-Range CR Scores	0	142,916
Duplicated Record	6	142,910

Table 5.12. Mathematics Grade 8 Data Cleaning

Exclusion Rule	# Deleted	# Cases Remain
Initial Number of Cases	n/a	159,467
Wrong Subject	0	159,467
No Grade	0	159,467
Wrong Grade	46	159,421
Language or Mismatched Form	2,498	156,923
School Type	182	156,741
Missing Entire Book	49,062	107,679
Invalid Score	3	107,676
Out-of-Range CR Scores	0	107,676
Duplicated Record	6	107,670

5.3. Classical Analysis and Calibration Sample Characteristics

The cleaned and sampled-down data sets included more than 98% of New York State students and were used for classical analyses, calibration, and linking. The demographic characteristics of students in these data sets are presented in Tables 5.13–5.18 and Tables 5.19–5.24 for ELA and Mathematics, respectively. The Needs/Resource Capacity Category (NRC) is assigned at the district level and is an indicator of district and school socioeconomic status. The ethnicity and gender designations are based on student-level information.

Table 5.13. ELA Grade 3 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	86,944	49.65
	Male	88,154	50.35
Ethnicity	Asian	17,794	10.16
	Black	31,280	17.86
	Hispanic	48,492	27.69
	American Indian	1,152	0.66
	Multiracial	5,122	2.93
	Pacific Islander	488	0.28
	White	70,770	40.42
NRC	New York	64,712	36.96
	Big 4 Cities	7,726	4.41
	Urban/Suburban	14,236	8.13
	High Needs Rural	9,868	5.64
	Average Needs	40,722	23.26
	Low Needs	17,939	10.25
	Charter School	11,164	6.38

Demographic Category		N-Count	% of Total N-Count
	Religious and Independent	8,731	4.99
SWD	No	152,749	87.24
	Yes	22,349	12.76
SUA	No	155,829	89.00
	Yes	19,269	11.00
ELL	No	156,567	89.42
	Yes	18,531	10.58

*The total n-count was 175,098.

Table 5.14. ELA Grade 4 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	87,697	50.16
	Male	87,124	49.84
Ethnicity	Asian	18,399	10.52
	Black	31,452	17.99
	Hispanic	48,380	27.67
	American Indian	1,170	0.67
	Multiracial	4,542	2.60
	Pacific Islander	548	0.31
	White	70,330	40.23
NRC	New York	65,999	37.75
	Big 4 Cities	7,537	4.31
	Urban/Suburban	13,332	7.63
	High Needs Rural	9,548	5.46
	Average Needs	38,875	22.24
	Low Needs	17,686	10.12
	Charter School	9,500	5.43
	Religious and Independent	12,344	7.06
SWD	No	151,722	86.79
	Yes	23,099	13.21
SUA	No	153,379	87.73
	Yes	21,442	12.27
ELL	No	159,858	91.44
	Yes	14,963	8.56

*The total n-count was 174,821.

Table 5.15. ELA Grade 5 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	81,838	49.72
	Male	82,758	50.28
Ethnicity	Asian	17,365	10.55
	Black	30,504	18.53
	Hispanic	45,194	27.46
	American Indian	1,123	0.68
	Multiracial	3,882	2.36
	Pacific Islander	606	0.37
	White	65,922	40.05
NRC	New York	63,287	38.45
	Big 4 Cities	6,915	4.20
	Urban/Suburban	12,618	7.67
	High Needs Rural	9,115	5.54
	Average Needs	37,090	22.53
	Low Needs	17,300	10.51
	Charter School	9,815	5.96
	Religious and Independent	8,456	5.14
SWD	No	141,331	85.87
	Yes	23,265	14.13
SUA	No	142,485	86.57
	Yes	22,111	13.43
ELL	No	151,702	92.17
	Yes	12,894	7.83

*The total n-count was 164,596.

Table 5.16. ELA Grade 6 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	79,546	49.28
	Male	81,878	50.72
Ethnicity	Asian	17,207	10.66
	Black	31,028	19.22
	Hispanic	44,338	27.47
	American Indian	1,138	0.70
	Multiracial	3,280	2.03
	Pacific Islander	478	0.30
	White	63,955	39.62

Demographic Category		N-Count	% of Total N-Count
NRC	New York	62,289	38.59
	Big 4 Cities	6,534	4.05
	Urban/Suburban	11,343	7.03
	High Needs Rural	8,250	5.11
	Average Needs	34,490	21.37
	Low Needs	16,496	10.22
	Charter School	10,412	6.45
	Religious and Independent	11,610	7.19
SWD	No	137,697	85.30
	Yes	23,727	14.70
SUA	No	139,552	86.45
	Yes	21,872	13.55
ELL	No	149,871	92.84
	Yes	11,553	7.16

*The total n-count was 161,424.

Table 5.17. ELA Grade 7 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	74,881	49.15
	Male	77,457	50.85
Ethnicity	Asian	17,121	11.24
	Black	29,097	19.10
	Hispanic	40,403	26.52
	American Indian	1,110	0.73
	Multiracial	2,845	1.87
	Pacific Islander	435	0.29
	White	61,327	40.26
NRC	New York	61,134	40.13
	Big 4 Cities	6,006	3.94
	Urban/Suburban	10,238	6.72
	High Needs Rural	7,926	5.20
	Average Needs	31,261	20.52
	Low Needs	17,065	11.20
	Charter School	9,977	6.55
	Religious and Independent	8,731	5.73
SWD	No	129,975	85.32

Demographic Category		N-Count	% of Total N-Count
SUA	Yes	22,363	14.68
	No	131,722	86.47
	Yes	20,616	13.53
ELL	No	141,443	92.85
	Yes	10,895	7.15

*The total n-count was 152,338.

Table 5.18. ELA Grade 8 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	69,786	48.73
	Male	73,421	51.27
Ethnicity	Asian	16,568	11.57
	Black	29,194	20.39
	Hispanic	39,307	27.45
	American Indian	1,139	0.80
	Multiracial	2,129	1.49
	Pacific Islander	435	0.30
	White	54,435	38.01
NRC	New York	60,999	42.59
	Big 4 Cities	5,963	4.16
	Urban/Suburban	9,352	6.53
	High Needs Rural	7,355	5.14
	Average Needs	27,774	19.39
	Low Needs	14,394	10.05
	Charter School	8,191	5.72
	Religious and Independent	9,179	6.41
SWD	No	122,728	85.70
	Yes	20,479	14.30
SUA	No	124,169	86.71
	Yes	19,038	13.29
ELL	No	133,774	93.41
	Yes	9,433	6.59

*The total n-count was 143,207.

Table 5.19. Mathematics Grade 3 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	87,363	49.06

Demographic Category		N-Count	% of Total N-Count
	Male	90,723	50.94
Ethnicity	Asian	18,619	10.46
	Black	31,919	17.92
	Hispanic	50,809	28.53
	American Indian	1,188	0.67
	Multiracial	4,784	2.69
	Pacific Islander	511	0.29
	White	70,256	39.45
NRC	New York	68,518	38.47
	Big 4 Cities	7,783	4.37
	Urban/Suburban	14,259	8.01
	High Needs Rural	9,702	5.45
	Average Needs	40,506	22.75
	Low Needs	17,950	10.08
	Charter School	11,587	6.51
	Religious and Independent	7,781	4.37
SWD	No	153,477	86.18
	Yes	24,609	13.82
SUA	No	155,533	87.34
	Yes	22,553	12.66
ELL	No	156,940	88.13
	Yes	21,146	11.87

*The total n-count was 178,086.

Table 5.20. Mathematics Grade 4 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	87,324	49.43
	Male	89,355	50.57
Ethnicity	Asian	19,258	10.90
	Black	31,928	18.07
	Hispanic	50,248	28.44
	American Indian	1,190	0.67
	Multiracial	4,301	2.43
	Pacific Islander	569	0.32
	White	69,185	39.16
NRC	New York	69,786	39.50
	Big 4 Cities	7,534	4.26

Demographic Category		N-Count	% of Total N-Count
NRC	Urban/Suburban	13,256	7.50
	High Needs Rural	9,323	5.28
	Average Needs	38,564	21.83
	Low Needs	17,838	10.10
	Charter School	9,659	5.47
	Religious and Independent	10,719	6.07
SWD	No	151,585	85.80
	Yes	25,094	14.20
SUA	No	152,275	86.19
	Yes	24,404	13.81
ELL	No	159,361	90.20
	Yes	17,318	9.80

*The total n-count was 176,679.

Table 5.21. Mathematics Grade 5 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	82,030	49.25
	Male	84,540	50.75
Ethnicity	Asian	18,043	10.83
	Black	30,690	18.42
	Hispanic	46,789	28.09
	American Indian	1,141	0.68
	Multiracial	3,619	2.17
	Pacific Islander	623	0.37
	White	65,665	39.42
NRC	New York	66,795	40.10
	Big 4 Cities	6,940	4.17
	Urban/Suburban	12,425	7.46
	High Needs Rural	8,770	5.27
	Average Needs	36,368	21.83
	Low Needs	17,260	10.36
	Charter School	9,860	5.92
	Religious and Independent	8,152	4.89
SWD	No	141,644	85.04
	Yes	24,926	14.96
SUA	No	141,679	85.06

Demographic Category		N-Count	% of Total N-Count
	Yes	24,891	14.94
ELL	No	151,346	90.86
	Yes	15,224	9.14

*The total n-count was 166,570.

Table 5.22. Mathematics Grade 6 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	79,222	48.91
	Male	82,748	51.09
Ethnicity	Asian	17,755	10.96
	Black	30,706	18.96
	Hispanic	44,907	27.73
	American Indian	1,136	0.70
	Multiracial	3,044	1.88
	Pacific Islander	492	0.30
	White	63,930	39.47
NRC	New York	64,195	39.63
	Big 4 Cities	6,498	4.01
	Urban/Suburban	10,914	6.74
	High Needs Rural	7,902	4.88
	Average Needs	33,550	20.71
	Low Needs	16,309	10.07
	Charter School	10,518	6.49
	Religious and Independent	12,084	7.46
SWD	No	138,165	85.30
	Yes	23,805	14.70
SUA	No	139,011	85.83
	Yes	22,959	14.17
ELL	No	148,150	91.47
	Yes	13,820	8.53

*The total n-count was 161,970.

Table 5.23. Mathematics Grade 7 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	69,674	48.75
	Male	73,236	51.25
Ethnicity	Asian	17,079	11.95
	Black	27,800	19.45
	Hispanic	39,562	27.68
	American Indian	1,100	0.77
	Multiracial	2,251	1.58
	Pacific Islander	429	0.30
	White	54,689	38.27

Demographic Category		N-Count	% of Total N-Count
NRC	New York	62,622	43.82
	Big 4 Cities	6,003	4.20
	Urban/Suburban	8,551	5.98
	High Needs Rural	7,066	4.94
	Average Needs	26,242	18.36
	Low Needs	15,989	11.19
	Charter School	9,930	6.95
	Religious and Independent	6,507	4.55
SWD	No	121,403	84.95
	Yes	21,507	15.05
SUA	No	122,351	85.61
	Yes	20,559	14.39
ELL	No	130,836	91.55
	Yes	12,074	8.45

*The total n-count was 142,910.

Table 5.24. Mathematics Grade 8 Sample Characteristics

Demographic Category		N-Count	% of Total N-Count
Gender	Female	51,545	47.87
	Male	56,125	52.13
Ethnicity	Asian	11,110	10.32
	Black	23,673	21.99
	Hispanic	33,077	30.72
	American Indian	901	0.84
	Multiracial	1,374	1.28
	Pacific Islander	345	0.32
	White	37,190	34.54
NRC	New York	51,257	47.61
	Big 4 Cities	5,302	4.92
	Urban/Suburban	6,359	5.91
	High Needs Rural	5,411	5.03
	Average Needs	16,047	14.90
	Low Needs	7,562	7.02
	Charter School	6,398	5.94
	Religious and Independent	9,334	8.67
SWD	No	89,620	83.24

Demographic Category		N-Count	% of Total N-Count
SUA	Yes	18,050	16.76
	No	90,207	83.78
	Yes	17,463	16.22
ELL	No	96,905	90.00
	Yes	10,765	10.00

*The total n-count was 107,670.

5.4. Classical Data Analysis

Classical data analysis of the NYSTP Grades 3–8 ELA and Mathematics Tests consists of several important elements. One element is the analysis of item-level statistical information about student performance. It is important to verify that the items and test forms function as intended. If any serious error were to occur with an item (e.g., a printing error or two correct answers to one item), item analysis is the stage at which errors should be flagged and evaluated for rectification (suppression, credit, or other acceptable solution). Analyses of test-level data comprise the second element of classical data analysis. These include examination of the raw score (RS) statistics (mean and standard deviation or “SD”) and test reliability measures Cronbach’s alpha (Cronbach, 1951) and Feldt-Raju coefficient (Qualls, 1995). Additionally, classical DIF analysis is conducted at this stage. DIF analysis includes computation of standardized mean differences and Mantel-Haenszel statistics for New York State items to identify potential item bias. All classical data analysis results contribute information on the validity and reliability of the tests (see also Section 3, “Validity,” and Section 7, “Reliability and Standard Error of Measurement”).

5.4.1. Item Difficulty and Point Biserial Correlation Coefficients

Item difficulty is classically measured by the p -value statistic. It assesses the proportion of students who responded correctly to each MC item or the average proportion of the maximum score that students earned on each CR item. It is important to have a good range of p -values to increase test information and to avoid floor or ceiling effects. P -values represent the overall degree of difficulty, but do not account for demonstrated student performance on other test items. Usually, p -value information is coupled with point biserial (pbis) statistics, to verify that items are functioning as intended. In Appendix M, Tables M1–M12 illustrate classical test statistics for all items on each grade-level test. Appendix F provides general psychometric guidelines for operational item selection.

Item difficulties (p -values) for the ELA tests ranged from 0.33 to 0.92. For Grade 3, the item p -values ranged from 0.37 to 0.84, with a mean of 0.57. For Grade 4, the item p -values ranged from 0.33 to 0.81, with a mean of 0.57. For Grade 5, the item p -values ranged from 0.39 to 0.89, with a mean of 0.59. For Grade 6, the item p -values ranged from 0.42 to 0.84, with a mean of 0.63. For Grade 7, the item p -values ranged from 0.41 to 0.89, with a mean of 0.62. For Grade 8, the item p -values ranged from 0.44 to 0.92, with a mean of 0.67. These p -value statistics are in Appendix M, Tables M1–M6, along with other classical test statistics of the keys.

Item difficulties (p -values) on the Mathematics tests ranged from 0.15 to 0.94. For Grade 3, the item p -values ranged from 0.30 to 0.94, with a mean of 0.65. For Grade 4, the item p -values

ranged from 0.32 to 0.93, with a mean of 0.64. For Grade 5, the item p -values ranged from 0.19 to 0.88, with a mean of 0.60. For Grade 6, the item p -values ranged from 0.21 to 0.85, with a mean of 0.53. For Grade 7, the item p -values ranged from 0.29 to 0.77, with a mean of 0.51. For Grade 8, the item p -values ranged from 0.15 to 0.83, with a mean of 0.46. These statistics are provided in Appendix M, Tables M7–M12, along with other classical test statistics.

Point-biserial statistics are used to examine item-test correlations, or item discrimination, for MC items. The pbis correlation for the key (i.e., the correct answer) is a measure of internal consistency, while pbis for specific response options aid in flagging possible alternate keys; each is a correlation that ranges between $+/-1$. It is the correlation of students' responses to an item relative to their performance on the rest of the test and, unless otherwise noted, this discussion will be limited to the point biserial of the correct response with the remainder of the test.

Point-biserial correlations from the operational analyses are presented in Appendix M Tables M1–M12. The column labeled “Pbis Key” contains the point biserial correlation associated with the correct response. The guideline for building the NYSTP Grades 3–8 ELA and Mathematics Tests was that the point-biserial correlation (from the field test analyses as item statistics from field testing were only available during form building) for the key for MC items should be equal to or greater than .20, which would indicate that students who responded correctly to that item also tended to do well on the overall test. The few exceptions to this guideline were due to content considerations which required the inclusion of particular items. Decisions to use such items were made very carefully, and no item with a negative point-biserial correlation was allowed on the test.

Point biserials for correct answer options on the ELA tests ranged from 0.11 to 0.70, as shown in Appendix M, Tables M1–M6. For Grade 3, the item pbis values ranged from 0.27 to 0.70, with a mean of 0.45. For Grade 4, the item pbis values ranged from 0.16 to 0.69, with a mean of 0.40. For Grade 5, the item pbis values ranged from 0.16 to 0.69, with a mean of 0.37. For Grade 6, the item pbis values ranged from 0.22 to 0.68, with a mean of 0.43. For Grade 7, the item pbis values ranged from 0.14 to 0.68, with a mean of 0.41. For Grade 8, the item pbis values ranged from 0.11 to 0.69, with a mean of 0.41.

Point biserials for correct answer options on the Mathematics tests ranged from 0.21 to 0.77, as shown in Appendix M, Tables M7–M12. For Grade 3, the item pbis values ranged from 0.27 to 0.71, with a mean of 0.48. For Grade 4, the item pbis values ranged from 0.34 to 0.71, with a mean of 0.49. For Grade 5, the item pbis values ranged from 0.22 to 0.73, with a mean of 0.50. For Grade 6, the item pbis values ranged from 0.21 to 0.70, with a mean of 0.47. For Grade 7, the item pbis values ranged from 0.27 to 0.77, with a mean of 0.50. For Grade 8, the item pbis values ranged from 0.21 to 0.72, with a mean of 0.44.

5.4.2. *Omit Rates*

Omit rates (i.e., percentage of students not answering a given item) are routinely checked, based on test data, after each administration. Tables M1–M12 in Appendix M show the omit rates for items on the Grades 3–8 ELA and Mathematics Tests, respectively. The industry standard general rule of thumb is that omit rates for multiple-choice items should be less than 5%. Omit rates across multiple-choice and constructed-response items on the Grades 3–8 ELA and Mathematics Tests typically ranged from 0% to 3%. As may be expected, omit rates tended to

increase for items at the end of the test booklets. That is, omit rates remained within the acceptable range for large-scale achievement tests.

5.4.3. *Differential Item Functioning (DIF)*

Classical differential item functioning (DIF) analyses are statistical methods for identifying items that are estimated to have functioned differently for one group (i.e., the “focal” group) as compared with another group (i.e., the “reference” group). In other words, DIF analysis only *flags* items that may later be judged by content experts to exhibit bias, rather than directly detecting bias. First, the psychometric phenomenon of DIF was extensively investigated and experts’ judgments of bias collected when items were field-tested, which reduced the likelihood of including any differentially functioning items on the operational forms for 2017. Turning to the analysis of the 2017 operational data, as discussed in Section 3.2.3. Detection of Bias, items flagged for DIF do not necessarily indicate item bias. For example, DIF may be attributed to true group differences on the content measured by the item or Type I error, which refers to statistically flagging items that have no true DIF. Operational items flagged for DIF are given additional scrutiny by content specialists, above and beyond the existing rounds of reviews by New York State educators, and those content specialists make the final judgment as to whether or not an item is biased for or against the focal group.

DIF was evaluated using two methods, both of which involve checks on statistical and practical significance. First, the Mantel-Haenszel (MH) method is employed for MC items. This non-parametric DIF method partitions the sample of examinees into categories based on total raw test scores. It then compares the log-odds ratio of keyed responses for the focal and reference groups. In terms of statistical significance, the Mantel-Haenszel method has a critical value of 6.63 (degrees of freedom = 1 for MC items; $\alpha = .01$) and as far as practical significance is concerned, it is compared to its corresponding delta-value. Delta-values are a commonly used metric in testing that indicates the magnitude of DIF. Typically, delta-values above 1.50 are considered indicative of moderate DIF that should be examined more closely (Zwick, Donoghue, and Grima, 1993). Second, the standardized mean difference (SMD) was computed for CR items. The SMD statistic (Dorans, Schmitt, and Bleistein, 1992) compares the mean scores of reference and focal groups, after adjusting for proficiency differences. The SMD was also evaluated for statistical significance and, in terms of practical significance, a moderate amount of DIF, for or against the focal group, is represented by an SMD with an absolute value between 0.10 and 0.19, inclusive; a large amount of DIF is represented by an SMD with an absolute value of 0.20 or greater.

Classical DIF analyses were conducted on subgroups of the Needs/Resource Capacity Category (focal group: High Needs; reference group: Low Needs), gender (focal group: Female; reference group: Male), ethnicity (focal groups: Black, Hispanic, and Asian; reference group: White), English language learners (focal group: English language learners; reference group: Non-English language learners), and mode (focal group: PBT students; reference group: CBT students). The DIF analyses were conducted using all cases from the clean data sets. Table 5.25 and Table 5.26 show the numbers of cases for the subgroups for ELA and Mathematics, respectively.

Table 5.25. ELA Classical DIF Sample N-Counts

Grade	Ethnicity				Gender		Needs/Resource Capacity Category		English Language Learners		Mode	
	Black/ African American	Hispanic/ Latino	Asian American	White								
					Female	Male	High	Low	ELL	Non-ELL	CBT	PBT
3	31,280	48,492	17,794	70,770	86,944	88,154	96,542	58,661	18,531	156,567	3,959	171,139
4	31,452	48,380	18,399	70,330	87,697	87,124	96,416	56,561	14,963	159,858	2,784	172,037
5	30,504	45,194	17,365	65,922	81,838	82,758	91,935	54,390	12,894	151,702	2,534	162,062
6	31,028	44,338	17,207	63,955	79,546	81,878	88,416	50,986	11,553	149,871	2,368	159,056
7	29,097	40,403	17,121	61,327	74,881	77,457	85,304	48,326	10,895	141,443	3,184	149,154
8	29,194	39,307	16,568	54,435	69,786	73,421	83,669	42,168	9,433	133,774	1,962	141,245

Table 5.26. Mathematics Classical DIF Sample N-Counts

Grade	Ethnicity				Gender		Needs/Resource Capacity Category		English Language Learners		Mode	
	Black/ African American	Hispanic/ Latino	Asian American	White								
					Female	Male	High	Low	ELL	Non-ELL	CBT	PBT
3	31,919	50,809	18,619	70,256	87,363	90,723	100,262	58,456	21,146	156,940	2,536	175,550
4	31,928	50,248	19,258	69,185	87,324	89,355	99,899	56,402	17,318	159,361	1,535	175,144
5	30,690	46,789	18,043	65,665	82,030	84,540	94,930	53,628	15,224	151,346	1,639	164,931
6	30,706	44,907	17,755	63,930	79,222	82,748	89,509	49,859	13,820	148,150	2,225	159,745
7	27,800	39,562	17,079	54,689	69,674	73,236	84,242	42,231	12,074	130,836	2,073	140,837
8	23,673	33,077	11,110	37,190	51,545	56,125	68,329	23,609	10,765	96,905	957	106,713

Table 5.27 (ELA) and Table 5.28 (Mathematics) present the number of items flagged for DIF by either of the classical methods described earlier. Appendix N provides a detailed list of items flagged by either one or both of these classical DIF methods, including DIF direction and associated DIF statistics.

Table 5.27. ELA Items Flagged for DIF

Grade	Flagged Items
3	7
4	6
5	8
6	16
7	17
8	11

Table 5.28. Mathematics Items Flagged for DIF

Grade	Flagged Items
3	9
4	4
5	2
6	5
7	4
8	4

As discussed in Section 3: Validity, items showing statistically significant DIF (flagged as described above for MH statistics on MC items and SMD statistics for CR items) do not necessarily pose bias. The items flagged with DIF were examined further by the content experts; no signs of potential content-based issues were discovered. The items are possibly functioning differently statistically.

Section 6: IRT Calibration and Linking

6.1. IRT Models and Rationale for Use

IRT allows for comparisons between items and scale scores, even those from different test forms, by using a common scale for all items and examinees (i.e., as if there were a hypothetical test that contained items from all forms). The three-parameter logistic (3PL) model (Lord and Novick, 1968; Lord, 1980) was used to analyze item responses on the MC items. For analysis of the CR items, the two-parameter partial credit (2PPC) model (Muraki, 1992; Yen, 1993) was used.

IRT is a statistical methodology that takes into account the fact that not all test items are alike and that not all test items provide the same amount of information in determining how much a student knows or can do. Computer programs that implement IRT models use actual student data to estimate the characteristics of the items on a test, called “parameters.” The parameter estimation process is called “item calibration.”

IRT models typically vary according to the number of parameters estimated. For the New York State tests, three parameters are estimated: the discrimination parameter, the difficulty parameter(s), and, for MC items, the guessing parameter. The discrimination parameter is an index of how well an item differentiates between high-performing and low-performing students. An item that cannot be answered correctly by low-performing students, but can be answered correctly by high-performing students, will have a high-discrimination value. The difficulty parameter is an index of how easy or difficult an item is. The higher the difficulty parameter is, the harder the item is. The guessing parameter is the probability that a student with very low proficiency will answer the item correctly.

Because the characteristics of MC and CR items are different, two IRT models were used in item calibration. The three-parameter logistic (3PL) model was used in the analysis of MC items. In this model, the probability that a student with proficiency θ responds correctly to item i is

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + \exp[-1.7a_i(\theta - b_i)]},$$

where

a_i is the item discrimination, b_i is the item difficulty, and c_i is the probability of a correct response from a very low-scoring student.

For analysis of the CR items, the 2PPC model was used. The 2PPC model is a special case of Bock’s (1972) nominal model. Bock’s model states that the probability of an examinee with proficiency θ having a score $(k - 1)$ at the k th level of the j th item is:

$$P_{jk}(\theta) = P(x_j = k - 1 | \theta) = \frac{\exp Z_{jk}}{\sum_{i=1}^{m_j} \exp Z_{ji}}, \quad k = 1 \text{ K } m_j,$$

where

$$Z_{jk} = A_{jk}\theta + C_{jk},$$

and

k is the item response category ($k = 1, 2, \dots, m_j$).

The m_j denotes the number of score levels for the j th item, and, typically, the highest score level is assigned $(m_j - 1)$ score points. For the special case of the 2PPC model used here, the following constraints were used:

$$A_{jk} = \alpha_j(k-1),$$

and

$$C_{jk} = -\sum_{i=0}^{k-1} \gamma_{ji},$$

where

$$\gamma_{j0} = 0,$$

and

α_j and γ_{ji} are the free parameters to be estimated from the data.

Each item has $(m_j - 1)$ independent γ_{ji} parameters and one α_j parameter; a total of m_j parameters are estimated for each item.

6.2. Calibration Sample

The cleaned data were used for calibration and linking of the NYSTP 2017 Grades 3–8 ELA and Mathematics Tests. It should be noted that the sample sizes were adequate, as the calibration and linking were performed using nearly all (96–99%, depending on grade level) of the New York State public and religious and independent school student population data in each tested grade. As shown in Tables 6.1–6.3 and Tables 6.4–6.6 for ELA and Mathematics, respectively, the 2017 operational test samples were generally comparable to 2016 populations in terms of NRC, student race and ethnicity, proportions of ELLs, proportions of students with disabilities, and proportions of students using testing accommodations.

Table 6.1. ELA Grades 3 and 4 Demographic Statistics

		Grade 3		Grade 4	
		2016	2017	2016	2017
Demographic Category		Population	Sample	Population	Sample
Gender	Female	49.51	49.65	49.32	50.16
	Male	50.49	50.35	50.68	49.84
Ethnicity	Asian	10.11	10.16	10.03	10.52
	Black	18.36	17.86	18.74	17.99
	Hispanic	28.41	27.69	27.89	27.67
	American Indian	0.69	0.66	0.63	0.67
	Multiracial	2.48	2.93	2.15	2.60

Demographic Category		Grade 3		Grade 4	
		2016	2017	2016	2017
		Population	Sample	Population	Sample
Ethnicity	Pacific Islander	0.32	0.28	0.37	0.31
	White	39.62	40.42	40.18	40.23
NRC	New York	39.42	36.96	39.22	37.75
	Big 4 Cities	4.31	4.41	4.17	4.31
	Urban/Suburban	7.73	8.13	7.46	7.63
	High Needs	5.36	5.64	5.18	5.46
	Rural	22.23	23.26	21.46	22.24
	Average Needs	9.74	10.25	9.60	10.12
	Low Needs	5.70	6.38	4.91	5.43
	Charter School	5.51	4.99	7.99	7.06
SWD	No	85.08	87.24	84.41	86.79
	Yes	14.92	12.76	15.59	13.21
SUA	No	93.22	89.00	92.28	87.73
	Yes	6.78	11.00	7.72	12.27
ELL	No	90.65	89.42	91.46	91.44
	Yes	9.35	10.58	8.54	8.56

Table 6.2. ELA Grades 5 and 6 Demographic Statistics

Demographic Category		Grade 5		Grade 6	
		2016	2017	2016	2017
		Population	Sample	Population	Sample
Gender	Female	49.06	49.72	49.07	49.28
	Male	50.94	50.28	50.93	50.72
Ethnicity	Asian	10.20	10.55	10.57	10.66
	Black	19.28	18.53	19.35	19.22
	Hispanic	27.82	27.46	26.88	27.47
	American Indian	0.67	0.68	0.68	0.70
	Multiracial	1.88	2.36	1.61	2.03
	Pacific Islander	0.28	0.37	0.27	0.30
	White	39.88	40.05	40.64	39.62
NRC	New York	40.40	38.45	38.52	38.59
	Big 4 Cities	4.04	4.20	3.96	4.05
	Urban/Suburban	7.35	7.67	6.66	7.03

Demographic Category		Grade 5		Grade 6	
		2016	2017	2016	2017
		Population	Sample	Population	Sample
NRC	High Needs Rural	5.13	5.54	4.99	5.11
	Average Needs	21.68	22.53	21.13	21.37
	Low Needs	10.11	10.51	10.34	10.22
	Charter School	5.59	5.96	6.32	6.45
	Religious and Independent	5.71	5.14	8.09	7.19
SWD	No	83.19	85.87	83.64	85.30
	Yes	16.81	14.13	16.36	14.70
SUA	No	91.59	86.57	91.62	86.45
	Yes	8.41	13.43	8.38	13.55
ELL	No	92.65	92.17	92.65	92.84
	Yes	7.35	7.83	7.35	7.16

Table 6.3. ELA Grades 7 and 8 Demographic Statistics

Demographic Category		Grade 7		Grade 8	
		2016	2017	2016	2017
		Population	Sample	Population	Sample
Gender	Female	48.72	49.15	48.61	48.73
	Male	51.28	50.85	51.39	51.27
Ethnicity	Asian	10.62	11.24	10.83	11.57
	Black	19.98	19.10	21.10	20.39
	Hispanic	27.02	26.52	27.44	27.45
	American Indian	0.73	0.73	0.66	0.80
	Multiracial	1.37	1.87	1.15	1.49
	Pacific Islander	0.28	0.29	0.26	0.30
	White	40.00	40.26	38.56	38.01
NRC	New York	41.35	40.13	42.79	42.59
	Big 4 Cities	3.99	3.94	3.95	4.16
	Urban/Suburban	6.68	6.72	6.37	6.53
	High Needs	5.07	5.20	4.94	5.14
	Average Needs	20.46	20.52	19.08	19.39
	Low Needs	10.64	11.20	10.02	10.05
	Charter School	5.70	6.55	4.94	5.72
	Religious and Independent	6.11	5.73	7.91	6.41

Demographic Category		Grade 7		Grade 8	
		2016	2017	2016	2017
		Population	Sample	Population	Sample
SWD	No	83.63	85.32	84.11	85.70
	Yes	16.37	14.68	15.89	14.30
SUA	No	92.11	86.47	92.37	86.71
	Yes	7.89	13.53	7.63	13.29
ELL	No	93.19	92.85	93.03	93.41
	Yes	6.81	7.15	6.97	6.59

Table 6.4. Mathematics Grades 3 and 4 Demographic Statistics

Demographic Category		Grade 3		Grade 4	
		2016	2017	2016	2017
		Population	Sample	Population	Sample
Gender	Female	49.36	49.06	49.21	49.43
	Male	50.64	50.94	50.79	50.57
Ethnicity	Asian	10.42	10.46	10.34	10.90
	Black	18.26	17.92	18.64	18.07
	Hispanic	28.64	28.53	28.18	28.44
	American Indian	0.69	0.67	0.63	0.67
	Multiracial	2.42	2.69	2.09	2.43
	Pacific Islander	0.32	0.29	0.38	0.32
	White	39.24	39.45	39.74	39.16
NRC	New York	40.06	38.47	39.92	39.50
	Big 4 Cities	4.36	4.37	4.19	4.26
	Urban/Suburban	7.67	8.01	7.33	7.50
	High Needs	5.25	5.45	5.06	5.28
	Rural	21.73	22.75	21.03	21.83
	Average Needs	9.67	10.08	9.64	10.10
	Low Needs	5.69	6.51	4.93	5.47
	Charter School	5.57	4.37	7.90	6.07
SWD	No	85.14	86.18	84.52	85.80
	Yes	14.86	13.82	15.48	14.20
SUA	No	93.00	87.34	90.58	86.19
	Yes	7.00	12.66	9.42	13.81
ELL	No	89.53	88.13	90.34	90.20
	Yes	10.47	11.87	9.66	9.80

Table 6.5. Mathematics Grades 5 and 6 Demographic Statistics

		Grade 5		Grade 6	
		2016	2017	2016	2017
Demographic Category		Population	Sample	Population	Sample
Gender	Female	48.97	49.25	49.01	48.91
	Male	51.03	50.75	50.99	51.09
Ethnicity	Asian	10.54	10.83	10.99	10.96
	Black	19.14	18.42	19.28	18.96
	Hispanic	28.18	28.09	27.31	27.73
	American Indian	0.68	0.68	0.67	0.70
	Multiracial	1.83	2.17	1.55	1.88
	Pacific Islander	0.29	0.37	0.28	0.30
Ethnicity	White	39.35	39.42	39.93	39.47
NRC	New York	41.23	40.10	39.73	39.63
	Big 4 Cities	4.06	4.17	3.98	4.01
	Urban/Suburban	7.22	7.46	6.43	6.74
	High Needs	4.94	5.27	4.77	4.88
	Rural	21.06	21.83	20.26	20.71
	Average Needs	10.04	10.36	10.24	10.07
	Low Needs	5.62	5.92	6.39	6.49
	Charter School	5.83	4.89	8.20	7.46
SWD	No	83.41	85.04	83.99	85.30
	Yes	16.59	14.96	16.01	14.70
SUA	No	90.23	85.06	89.96	85.83
	Yes	9.77	14.94	10.04	14.17
ELL	No	91.45	90.86	91.45	91.47
	Yes	8.55	9.14	8.55	8.53

Table 6.6. Mathematics Grades 7 and 8 Demographic Statistics

		Grade 7		Grade 8	
		2016	2017	2016	2017
Demographic Category		Population	Sample	Population	Sample
Gender	Female	48.66	48.75	47.86	47.87
	Male	51.34	51.25	52.14	52.13
Ethnicity	Asian	11.03	11.95	9.56	10.32
	Black	19.91	19.45	22.97	21.99

Demographic Category		Grade 7		Grade 8	
		2016	2017	2016	2017
		Population	Sample	Population	Sample
	Hispanic	27.64	27.68	30.92	30.72
	American Indian	0.73	0.77	0.67	0.84
	Multiracial	1.29	1.58	1.04	1.28
	Pacific Islander	0.29	0.30	0.27	0.32
	White	39.11	38.27	34.58	34.54
NRC	New York	43.08	43.82	46.60	47.61
	Big 4 Cities	3.95	4.20	4.55	4.92
	Urban/Suburban	6.34	5.98	6.52	5.91
	High Needs	4.76	4.94	4.77	5.03
	Rural	19.30	18.36	15.62	14.90
	Average Needs	10.36	11.19	7.04	7.02
	Low Needs	5.82	6.95	5.17	5.94
NRC	Charter School	6.38	4.55	9.73	8.67
SWD	No	84.02	84.95	81.71	83.24
	Yes	15.98	15.05	18.29	16.76
SUA	No	91.11	85.61	89.44	83.78
	Yes	8.89	14.39	10.56	16.22
ELL	No	91.75	91.55	89.76	90.00
	Yes	8.25	8.45	10.24	10.00

6.2.1. Calibration Process

The item parameters were estimated using Scientific Software International (SSI) Inc.'s IRTPRO Version 2.1 (Cai, Thissen, & du Toit, 2011) package. MC and CR items were calibrated simultaneously, using marginal maximum likelihood procedures.

The calibration of NYSTP 2017 Grades 3–8 ELA and Mathematics Tests did not exhibit any test-level issues. The estimated parameters were on the original theta scale, and all of the items were well within the prescribed parameter ranges. For both the Grades 3–8 ELA and Mathematics Tests, all calibration estimation results were reasonable. Tables 6.7 and 6.8 present the summaries of the calibration results for ELA and Mathematics, respectively. Additional details, including individual item parameter estimates, may be found in Appendix O, in Tables O13–O24. The parameter estimates are expressed on the theta metric and are defined below:

- MC items:
 - *a*-parameter is a discrimination parameter
 - *b*-parameter is a difficulty parameter
 - *c*-parameter is a guessing parameter

- CR items:
 - *alpha* is a discrimination parameter
 - *step* is a difficulty parameter for category m_j

As described in Section 6: IRT Calibration and Linking, above in Section 6.1. IRT Models and Rationale for Use, m_j denotes the number of score levels for the j th item, and, typically, the highest score level is assigned $(m_j - 1)$ score points. For the 2PPC model there are $m_j - 1$ independent steps and one alpha, for a total of m_j independent parameters estimated for each item, while there is one *a*-parameter and one *b*-parameter per item in the 3PL model.

Table 6.7. ELA Calibration Results

Grade	Item-Level			Student-Level		
	Largest	Range of b-		N-Count	Theta Est.*	
	a-Parameter	Parameters			Mean	SD
3	1.198	-1.624	1.010	174,910	0.01	0.94
4	1.253	-1.425	2.175	174,747	0.01	0.93
5	1.130	-1.721	1.661	164,526	0.00	0.94
6	1.161	-1.422	1.103	161,424	-0.00	0.94
7	1.430	-1.916	1.557	152,338	0.00	0.94
8	1.504	-2.191	1.634	143,207	-0.00	0.94

*Maximum *a posteriori* (MAP) theta estimates.

Table 6.8. Mathematics Calibration Results

Grade	Item-Level			Student-Level		
	Largest	Range of b-		N-Count	Theta Est.*	
	a-Parameter	Parameters			Mean	SD
3	1.591	-2.288	0.901	178,086	0.00	0.93
4	1.865	-2.111	0.968	176,679	0.00	0.93
5	1.930	-2.132	1.320	166,447	0.01	0.92
6	1.869	-1.603	1.849	161,690	0.02	0.93
7	3.117	-0.636	1.273	142,562	0.04	0.90
8	2.920	-1.025	1.596	107,489	0.06	0.89

*Maximum *a posteriori* (MAP) theta estimates.

6.3. Item-Model Fit

Item fit statistics provide evidence of the appropriateness of using an item in the 3PL or 2PPC model. The Q_I procedure described by Yen (1981) was used to measure fit to the three-parameter model. Students are rank-ordered on the basis of $\hat{\theta}$ values and sorted into ten cells with 10% of the sample in each cell. For each item, the number of students in cell k who answered item i , N_{ik} , and the number of students in that cell who answered item i correctly, R_{ik} , were determined. The observed proportion in cell k passing item i , O_{ik} , is R_{ik}/N_{ik} . The fit index for item i is:

$$Q_{Ii} = \sum_{k=1}^{10} \frac{N_{ik} (O_{ik} - E_{ik})^2}{E_{ik} (1 - E_{ik})}$$

with:

$$E_{ik} = \frac{1}{N_{ik}} \sum_{j \in \text{cell } k}^{N_{ik}} P_i(\hat{\theta}_j)$$

A modification of this procedure was used to measure fit to the 2PPC model. For the 2PPC model, Q_{Ij} was assumed to have an approximate chi-square distribution with the following degrees of freedom (df):

$$df = I(m_j - 1) - m_j,$$

where I is the total number of cells (usually 10) and m_j is the possible number of score levels for item j .

To adjust for differences in degrees of freedom among items, Q_I was transformed to Z_{Q_I} where:

$$Z_{Q_I} = (Q_I - df) / (2df)^{1/2}$$

The value of Z increases with sample size, when all else is equal. To use this standardized statistic to flag items for potential poor fit, it has been a common practice to vary the critical value for Z as a function of sample size. For the tests that have large calibration sample sizes, the criterion $Z_{Q_I} \text{Crit}$ was used to flag items and was calculated using the expression

$$Z_{Q_I} \text{Crit} = \left(\frac{N}{1500} \right) * 4,$$

where N is the calibration sample size.

To compute the Q_I and related statistics, a stratified sampling procedure was implemented in a way that a representative sample with the size of approximately 70,000 students was drawn at each grade level. Items were considered to have poor fit if the value of the obtained Z_{Q_I} was greater than the value of Z_{Q_I} critical. If the obtained Z_{Q_I} was less than Z_{Q_I} critical, the items were rated as having acceptable fit. The fact that the majority of the items in the NYSTP 2017 Grades 3–8 ELA and Mathematics Tests demonstrated good model fit further supports the use of the chosen models. Item fit statistics are presented in Tables O1–O12 in Appendix O.

6.4. Local Independence

In using IRT models, one of the assumptions made is that the items are locally independent; that a student's response to one item is not dependent upon his or her response to another item. In other words, when a student's proficiency is accounted for, his or her response to each item is statistically independent.

One way to measure the statistical independence of items within a test is via the Q_3 statistic (Yen, 1984). This statistic was obtained by correlating differences between students' observed and expected responses for pairs of items after taking into account overall test performance. The Q_3 statistic for binary items was computed as

$$d_{ij} \equiv u_{ij} - P_j(\hat{\theta}_i)$$

where $\hat{\theta}_i$ is the estimated trait value (i.e., proficiency) for the i th examinee; u_{ij} is the observed probability for the i th examinee to get the j th item correct and P_j is estimated probability for the i th examinee to get the j th item correct, and

$$Q_{3,ij} = r(d_{ij}, d_{j'})$$

The generalization to items with multiple response categories uses

$$d_{ij} \equiv x_{ij} - E_{ij}$$

where

$$E_{ij} \equiv E(x|\hat{\theta}_i) = \sum_{k=1}^{m_j} kP_{jk}(\hat{\theta}_i)$$

If a substantial number of items in the test demonstrate local dependence, these items may need to be calibrated separately. All pairs of items with Q_3 values greater than 0.20 were classified as significant for local dependency. The maximum value for this index is 1.00. When item pairs are flagged by Q_3 , the content of the flagged items is examined to identify possible sources of the local dependence. The primary concern about locally dependent items is that they contribute less psychometric information about examinee proficiency than do locally independent items, and therefore inflate score reliability estimates.

The Q_3 statistics were examined for all unique pairs of ELA and mathematics items. Items that were found to be significant in local dependency vary, depending on the subject and grade: one pair of items was found in ELA Grade 8. When reviewing the results for Mathematics, one pair of items each exceeded a correlation of 0.20 in Mathematics Grades 4, 7, and 8. The magnitudes of these statistics were not sufficient to warrant further concern or action (with the Q_3 values being 0.27 for the ELA test and ranging from 0.23 to 0.28 for the Mathematics tests).

6.5. Linking and Scaling

The purpose of linking was to place the 2017 item parameters and proficiency estimates on the same scale as those in 2016. The following steps constitute the linking process for each subject and grade:

1. Operational items as well as non-scored (i.e., external) anchor items were calibrated in IRTPRO.
2. The 2017 item parameter estimates for all anchor items—both scored and non-scored—enabled the establishment of the linking relationship via a test characteristic curve (TCC) method (Stocking and Lord, 1983; implemented in STUIRT, Kim, & Kolen, 2004) to the 2016 theta scale, using the established 2016 item parameter estimates for those same items. Table 6.9 and Table 6.10 present the resulting linking coefficients. The following parameters were linked using the formulas below:

$$a_i^E = a_i^C / M_1^E,$$

$$b_i^E = M_1^E \cdot b_i^C + M_2^E, \text{ and}$$

$$d_{ij}^E = d_{ij}^C + \left[\left(a_i^C / M_1^E \right) \right] \cdot M_2^E,$$

where

M_1^E is defined as the multiplicative adjustment for linking and M_2^E is the additive adjustment for linking. The superscript “E” denotes linked item parameter estimates, while the superscript “C” denotes calibrated item parameter estimates.

Table 6.9. ELA Linking Coefficients

Grade	M_1^E	M_2^E
3	1.006	0.242
4	0.990	0.219
5	1.102	0.071
6	1.017	-0.004
7	0.947	0.251
8	0.988	0.227

Table 6.10. Mathematics Linking Coefficients

Grade	M_1^E	M_2^E
3	1.170	0.264
4	1.200	0.158
5	1.130	0.248

Grade	M_1^E	M_2^E
6	1.259	0.125
7	1.152	0.230
8	1.162	-0.240

3. A raw-score-to-theta conversion chart was produced using the test characteristic curve (TCC) method (Stocking and Lord, 1983; see Section 0Scoring Procedure for more details) and implemented in POLYEQUATE (Kolen & Cui, 2004). The theta estimates associated with the TCC method ($\hat{\theta}_{TCC}$) must be linked back to the underlying theta scale established in the prior year (Spring 2016), and are computed as follows:

$$\theta^E = (M_1^E \cdot \hat{\theta}_{TCC}) + M_2^E$$

4. The TCC method does not produce theta estimates for raw scores below chance level or above the perfect score (highest obtainable raw score). In addition, for the scores at the low and high ends of the scale, some raw scores tended to have large theta estimates (for example, -7.999). Typically, the first obtainable theta value on a test corresponds to a very extreme theta value. The following adjustment/interpolation was conducted:

For any linked theta estimates (θ^E) that are outside of the range of -2.5 to 3, at the lower end of the scale, 0.25 was subtracted from the preceding theta value that is within the range; at the higher end of the scale, 0.25 was added to the previous theta value that is within the range, thus resulting in an adjusted theta estimate (θ^A) for those extremes. See the table below for an example at the lower end of the scale. Such an adjustment helps contain the theta scale within a reasonable range, and is standard practice in testing.

Raw Score	θ^E	θ^A
6	-5.30263	-3.37458
7	-3.66491	-3.12458
8	-3.03055	-2.87458
9	-2.76782	-2.62458
10	-2.37458	-2.37458

5. Once theta values were either estimated or interpolated for all raw scores, the raw-score-to-theta relationship was applied to each student, yielding a theta estimate corresponding to his or her raw score.
6. The adjusted theta estimates (presented in Table 6.11 and Table 6.12) were then scaled using the established scaling coefficients from the prior year (Spring 2016) according to the following formula:

$$ScaleScore = (M_1^S \cdot \theta^A) + M_2^S,$$

where

M_1^S is defined as the multiplicative scaling coefficient, and M_2^S is the additive scaling coefficient. M_1^S and M_2^S are applied to a true score (i.e., the linked theta estimate) in order to obtain a scale score.

Table 6.11. ELA Scaling Coefficients

Grade	M_1^S	M_2^S
3	31.8145	301.4946
4	32.0356	300.7619
5	32.0160	300.9540
6	32.2585	300.6730
7	31.9257	300.8012
8	31.6273	300.9795

Table 6.12. Mathematics Scaling Coefficients

Grade	M_1^S	M_2^S
3	32.2491	299.8560
4	32.6982	300.1764
5	32.2199	300.6932
6	32.4213	300.3769
7	31.2289	301.1438
8	31.8685	301.1430

7. Scale scores range, approximately, from 100 to 400 across grades. The lowest and highest observed scale score (LOSS and HOSS, respectively) may vary by grade.
8. A series of anchor set stability checks were performed before finalizing the anchor set for each subject and grade; see Section 6.6. Anchor Set Evaluation, which follows this one.
9. For conditional standard error of measurement (CSEM), the scale scores (both estimated and interpolated) were used to compute the information function and CSEM.

Throughout this process, NYSED psychometricians have reviewed, and a senior scientist from HumRRO has independently verified, the results generated by Questar psychometricians.

6.6. Anchor Set Evaluation

In order to determine if each item from the anchor set performs similarly to when it was administered in the prior year, comparisons of individual item characteristic curves (ICCs) and item parameter estimates from the previous and current administrations were made. Initial comparisons included a graphical inspection of the linearity of relationships between linked item parameter estimates from the 2016 and 2017 administrations. These revealed approximately linear relationships as well as similarities in item functions, and therefore provided support for

the selected linking method used herein. Additional analyses of the correlations between linked item parameter estimates also provided evidence of strong linear relationships.

A formal process for validating the anchor set by using an objective criterion was used to determine if any items ought to be considered for removal from the anchor set. The linked item parameter estimates were used to calculate a weighted, squared deviation of the current ICC from the previous ICC, across the range of ability (i.e., theta, or θ) and under a hypothetical normal distribution for θ . For a given item i , that quantity, called “d squared,” is given by

$$d_i^2 = \sum_k \{ [\text{Pr}_{i,16}(\theta_k) - \text{Pr}_{i,15}(\theta_k)]^2 \cdot g(\theta_k) \},$$

where i indexes anchor items; k indexes quadrature points for θ ; $\text{Pr}_{i,16}(\cdot)$ is the probability of a correct response to item i under the current calibration, while $\text{Pr}_{i,15}(\cdot)$ is the same quantity under the previous calibration; and $g(\theta_k)$ are weights for the quadrature points.

Historically, and as recently as the 2015 operational linking, a fixed criterion on this metric ($d_i^2 \geq 0.05$) has been used for flagging items to be considered for removal from linking. The same approach and criterion were used for the linking of the 2017 operational forms to the 2016 scale score scale. This procedure minimizes the weighted squared differences between the two ICCs for each MC item: one based on 2016 item parameter estimates and the other on 2017 estimates. The differential item performance was evaluated by examining previous and current item parameters. The following steps were taken:

1. Before the iterative procedures start, the initial linking was performed, using all of the eligible anchor items as an anchor set, as described in Section 6.5: Linking and Scaling. The initial linking coefficients (M_1^E and M_2^E) were obtained through the Stocking-Lord method.
2. The following process was repeated for at least five iterations or until the largest $d_i^2 < 0.05$ is reached, whichever was greater:
 - a. For each anchor item, d_i^2 was calculated as a weighted sum of the squared deviations between the ICCs based on old (2016) and new (2017) parameter estimates at each quadrature point and assuming a normal theta distribution.
 - b. The item having the largest d_i^2 was identified and removed from the anchor set.
 - c. The linking procedures described in Section 6.5: Linking and Scaling were performed with the newly reduced anchor set.
 - d. New raw-score-to-scale-score tables were prepared as described in Section 6.8. Scoring Procedure.
3. Select the linking coefficients (M_1^E and M_2^E) associated with the iteration selected in step 2 above.

The items that are implicitly proposed for removal from the anchor set, based on the process described above, were summarized and evaluated. The only subject where items were proposed and ultimately approved for removal from the anchor set was mathematics, and one item each was removed from the anchor sets for Grades 5, 6, and 7.

6.7. Test Characteristic Curves

Test Characteristic Curves (TCCs) provide an overview of the tests in the IRT scale score metric. The 2017 TCCs were generated using final item parameters for all reporting test items administered in Spring 2017. TCCs are the summation of all the item characteristic curves (ICCs) for items that contribute to the scale score. Conditional standard error of measurement (CSEM) curves graphically show the amount of measurement error at different performance levels. The TCCs and CSEM curves are presented in Figures 6.1–6.24.

Figure 6.1. ELA Grade 3 TCC

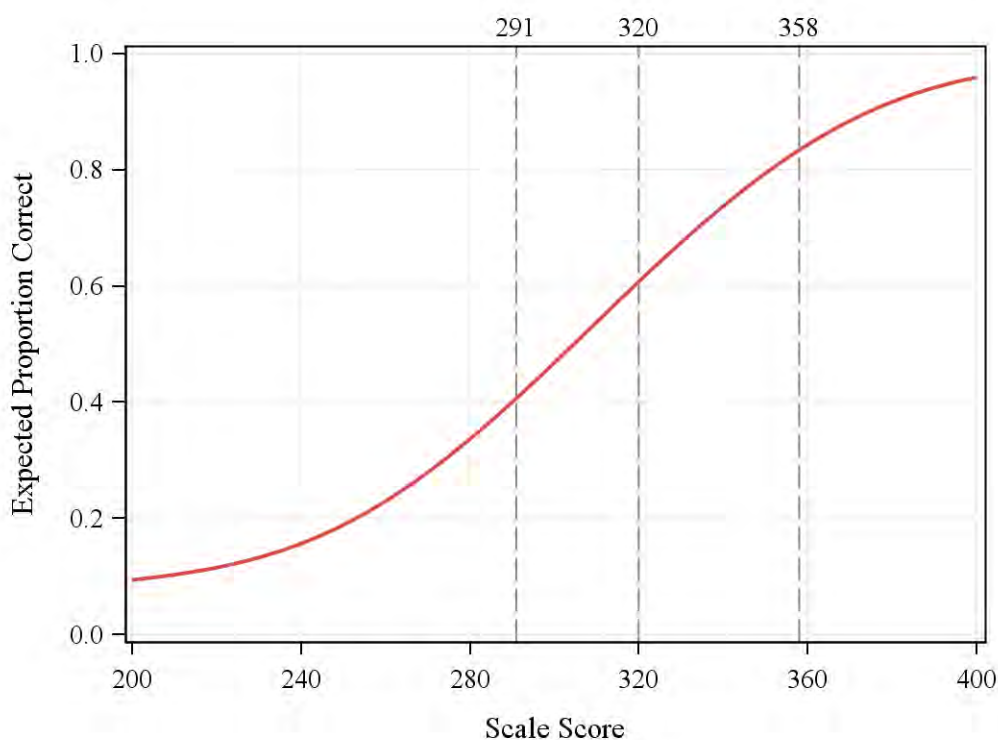


Figure 6.2. ELA Grade 3 CSEM Curve

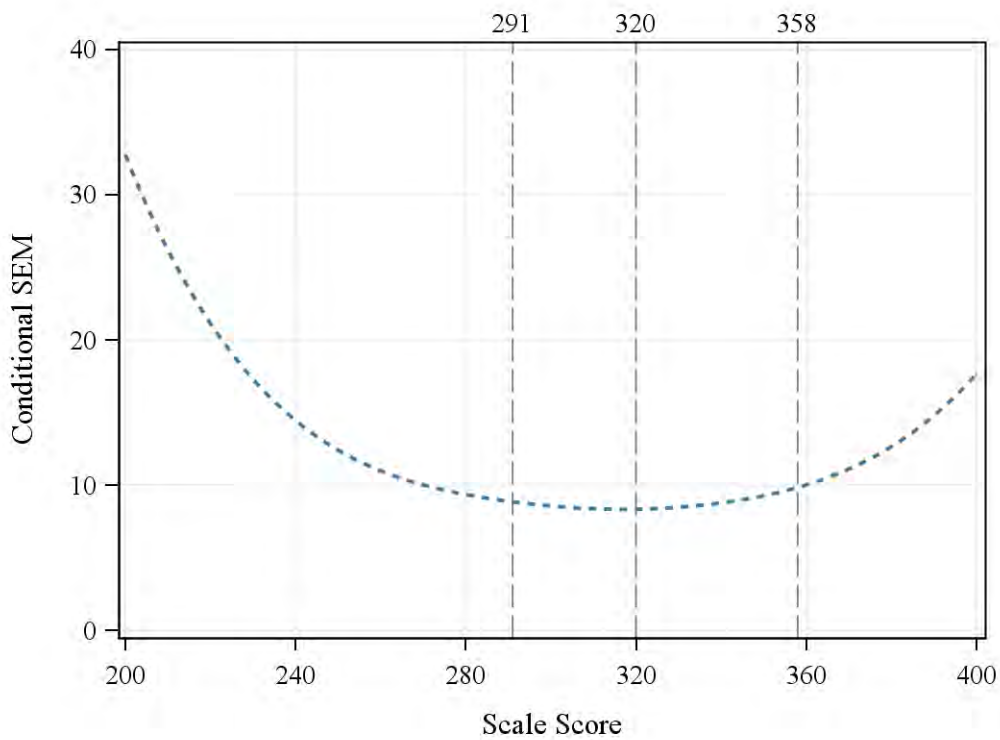


Figure 6.3. ELA Grade 4 TCC

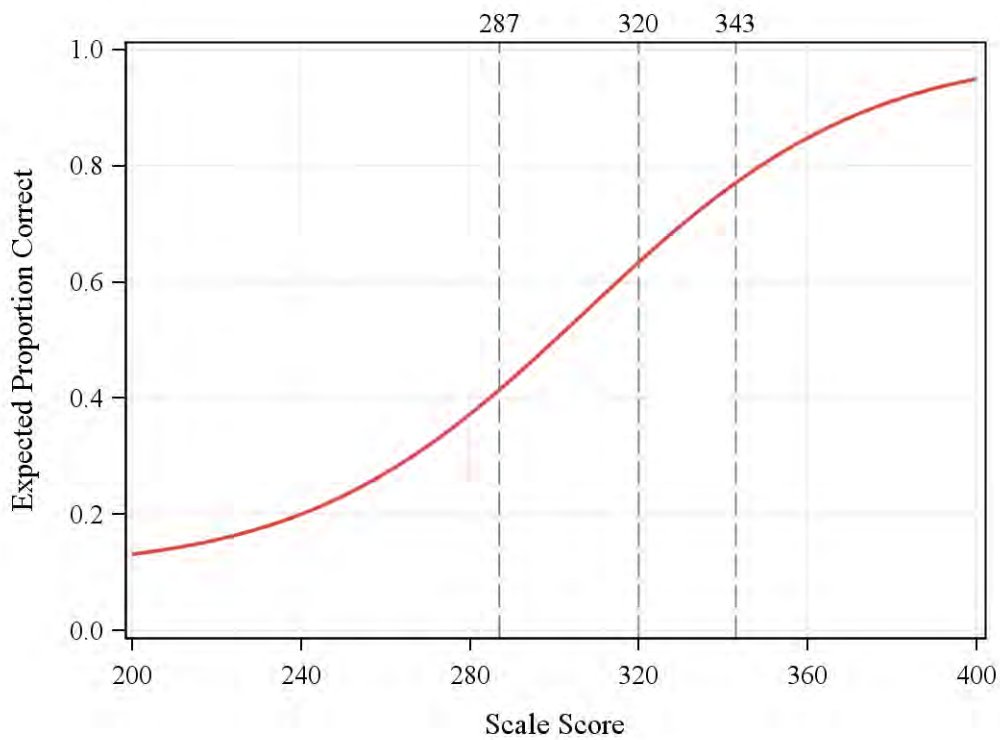


Figure 6.4. ELA Grade 4 CSEM Curve

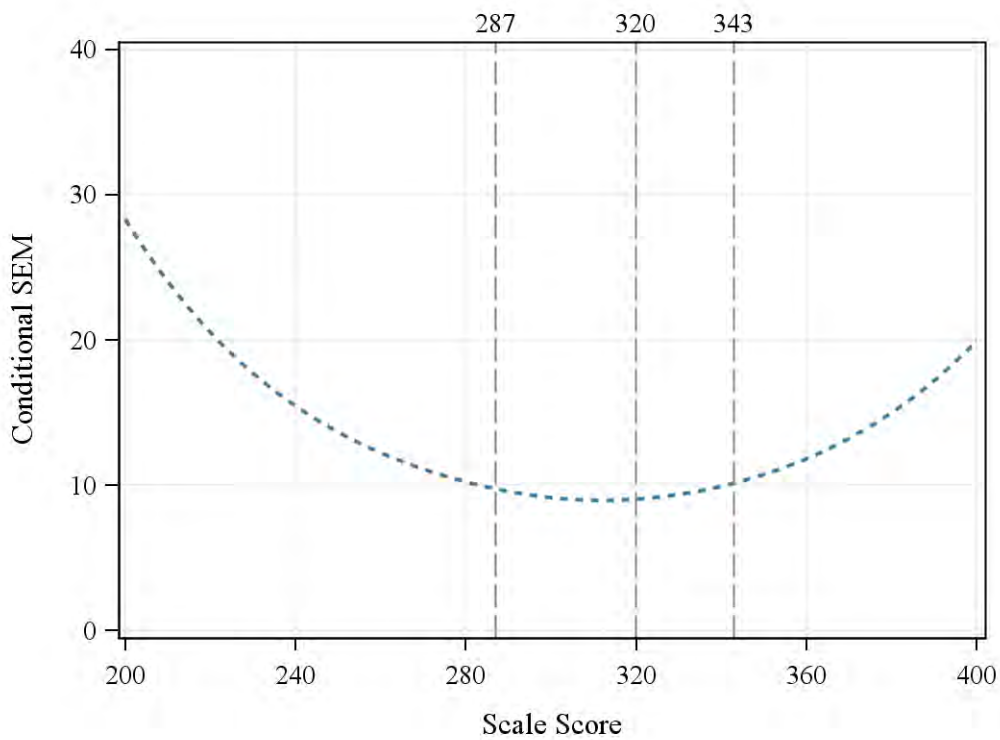


Figure 6.5. ELA Grade 5 TCC

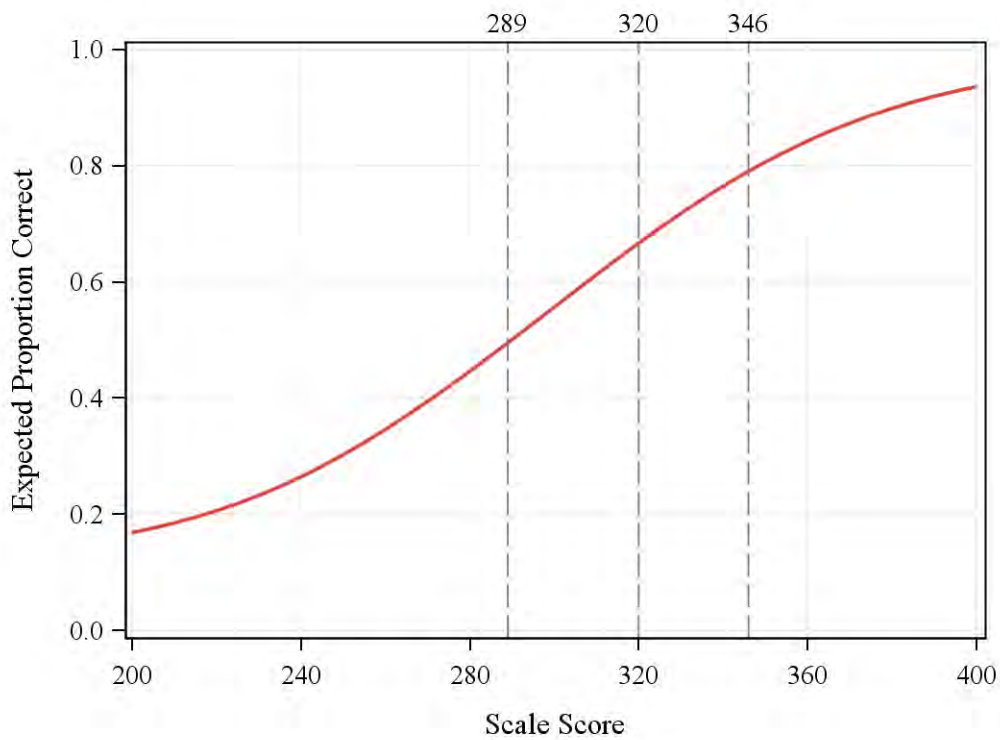


Figure 6.6. ELA Grade 5 CSEM Curve

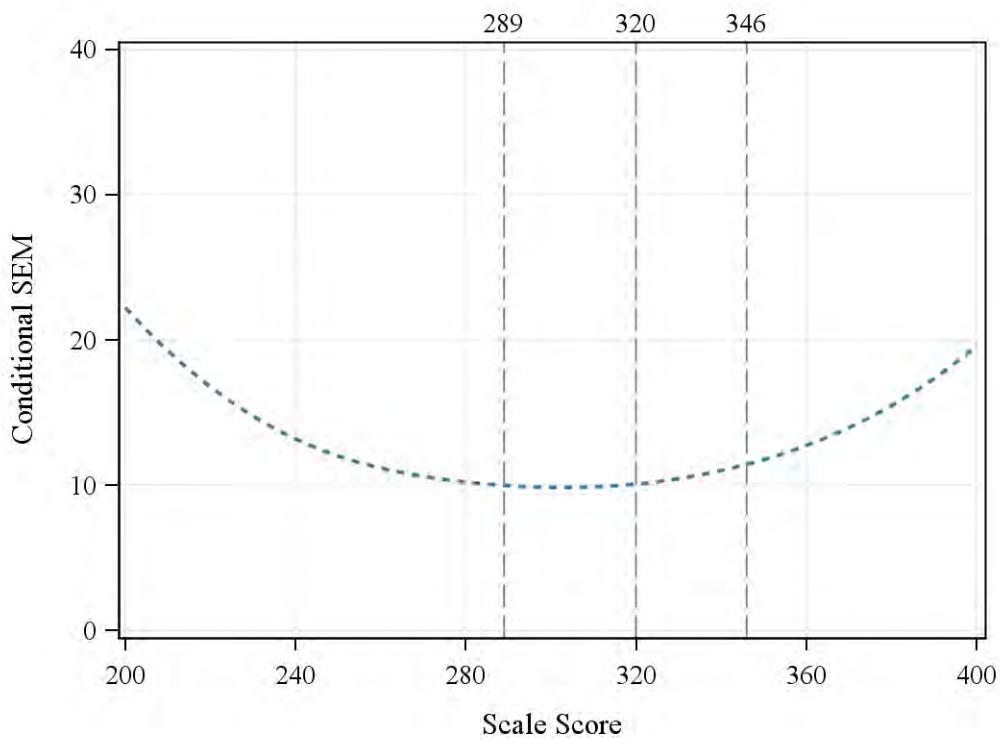


Figure 6.7. ELA Grade 6 TCC

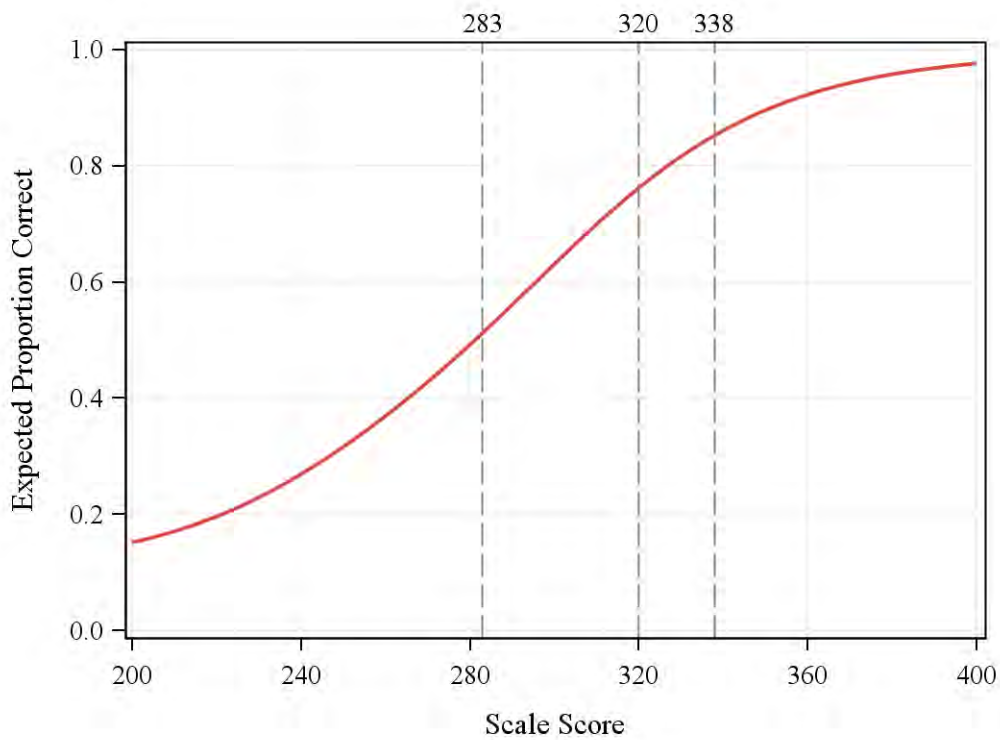


Figure 6.8. ELA Grade 6 CSEM Curve

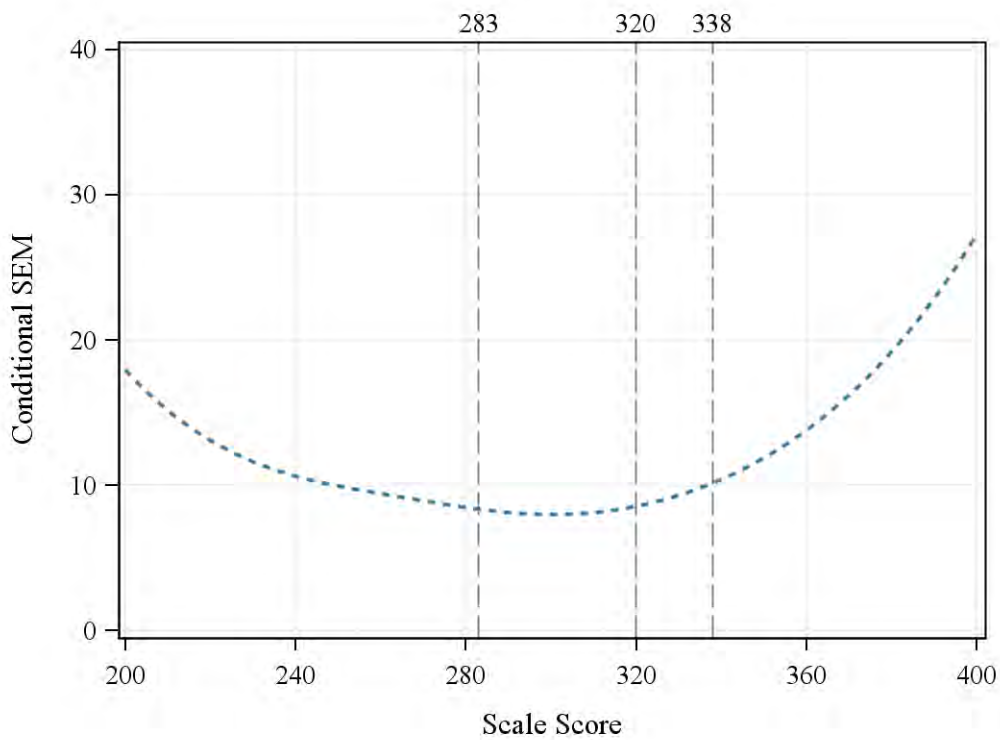


Figure 6.9. ELA Grade 7 TCC

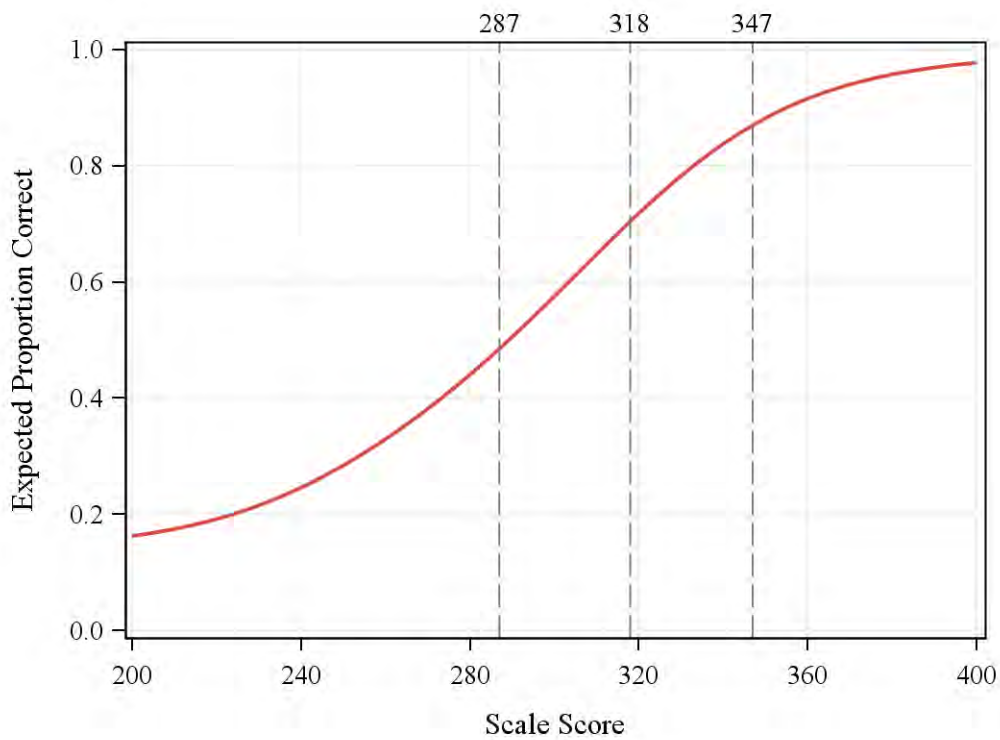


Figure 6.10. ELA Grade 7 CSEM Curve

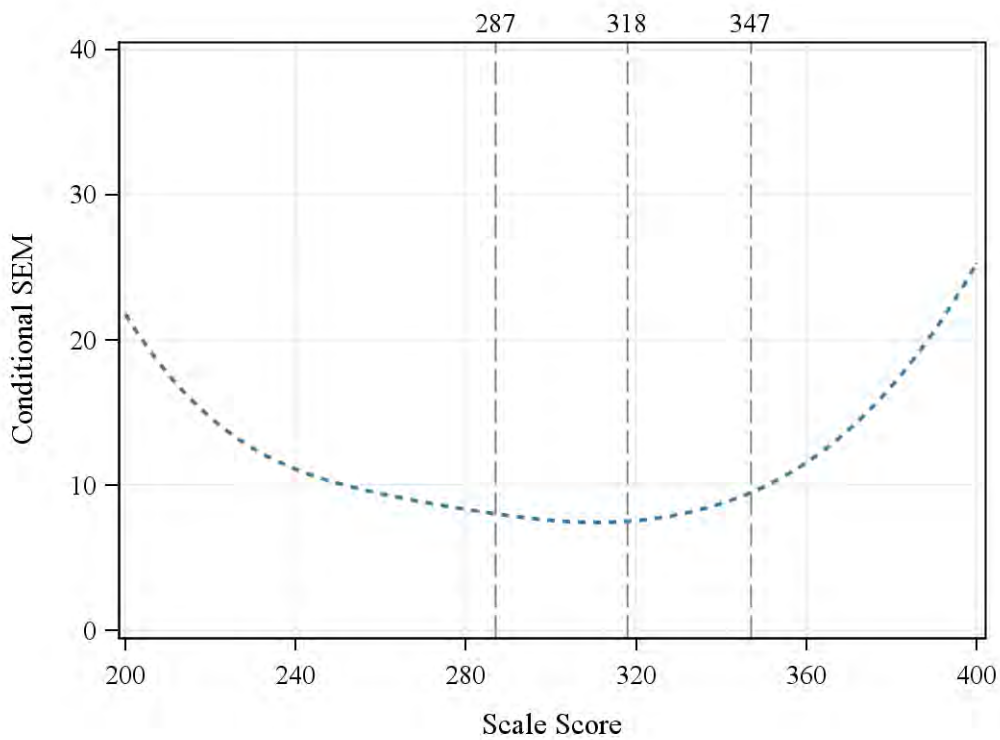


Figure 6.11. ELA Grade 8 TCC

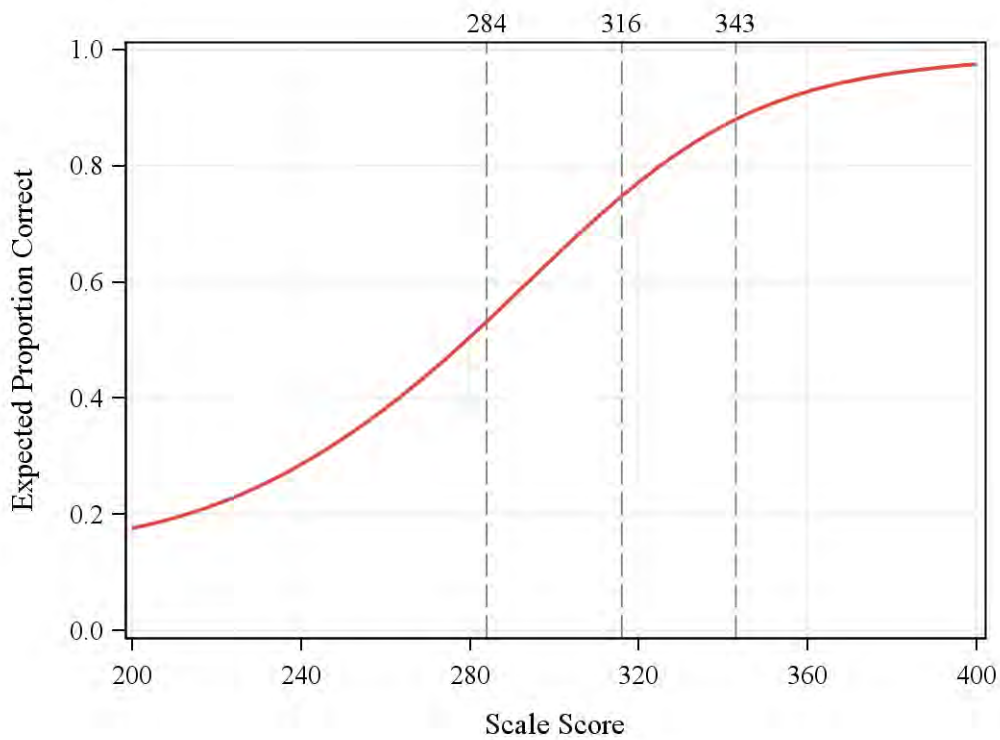


Figure 6.12. ELA Grade 8 CSEM Curve

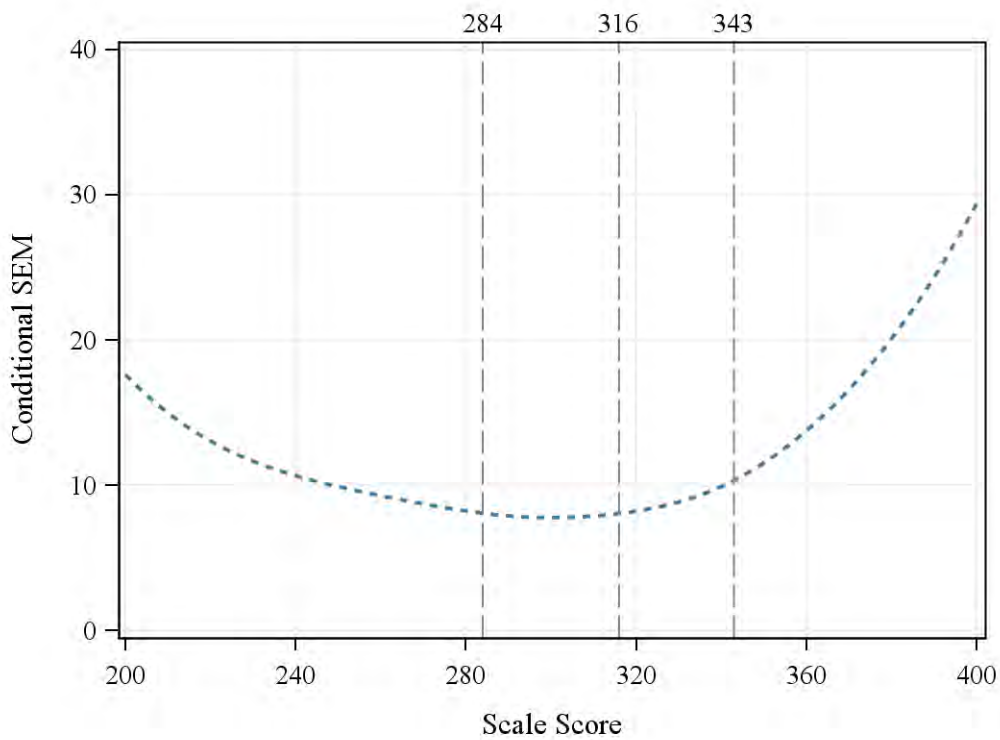


Figure 6.13. Mathematics Grade 3 TCC

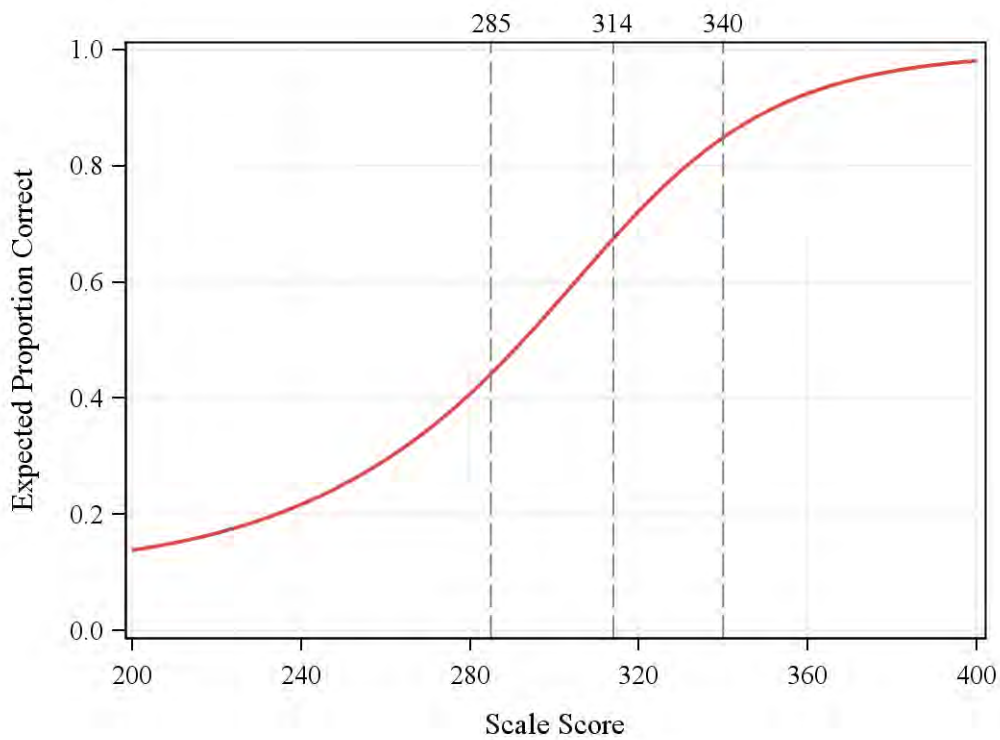


Figure 6.14. Mathematics Grade 3 CSEM Curve

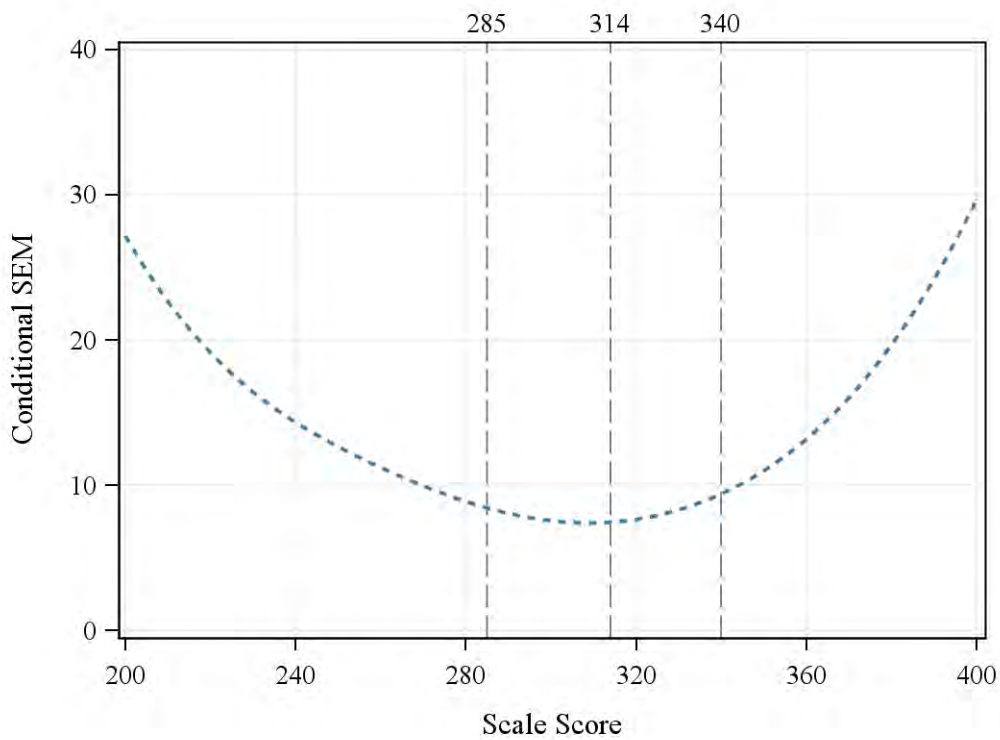


Figure 6.15. Mathematics Grade 4 TCC

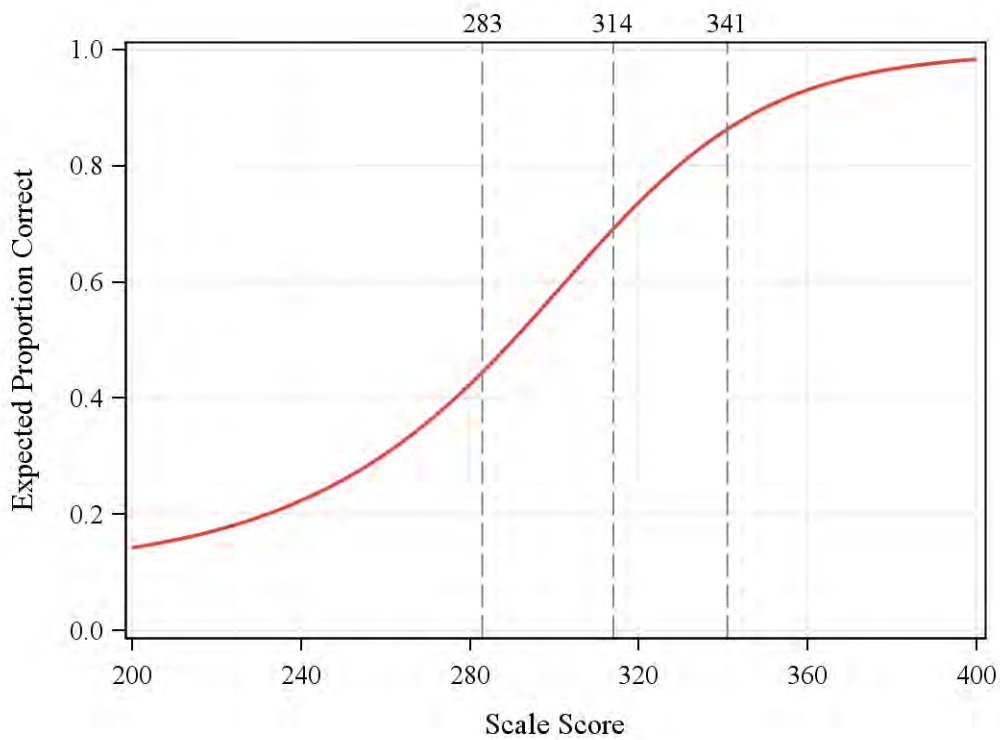


Figure 6.16. Mathematics Grade 4 CSEM Curve

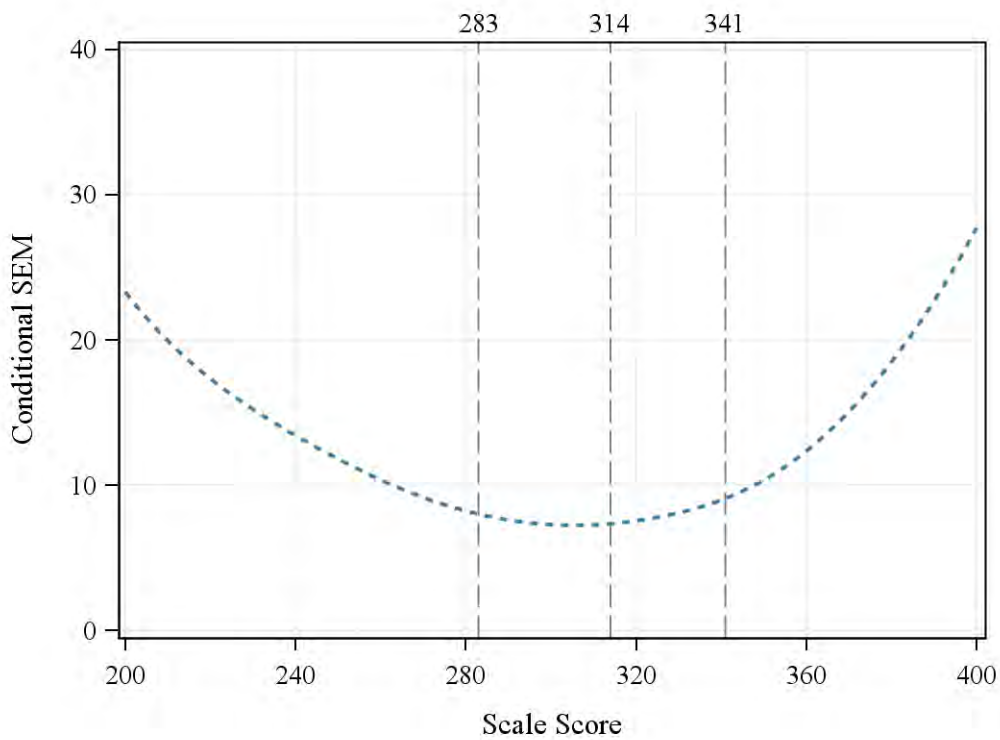


Figure 6.17. Mathematics Grade 5 TCC

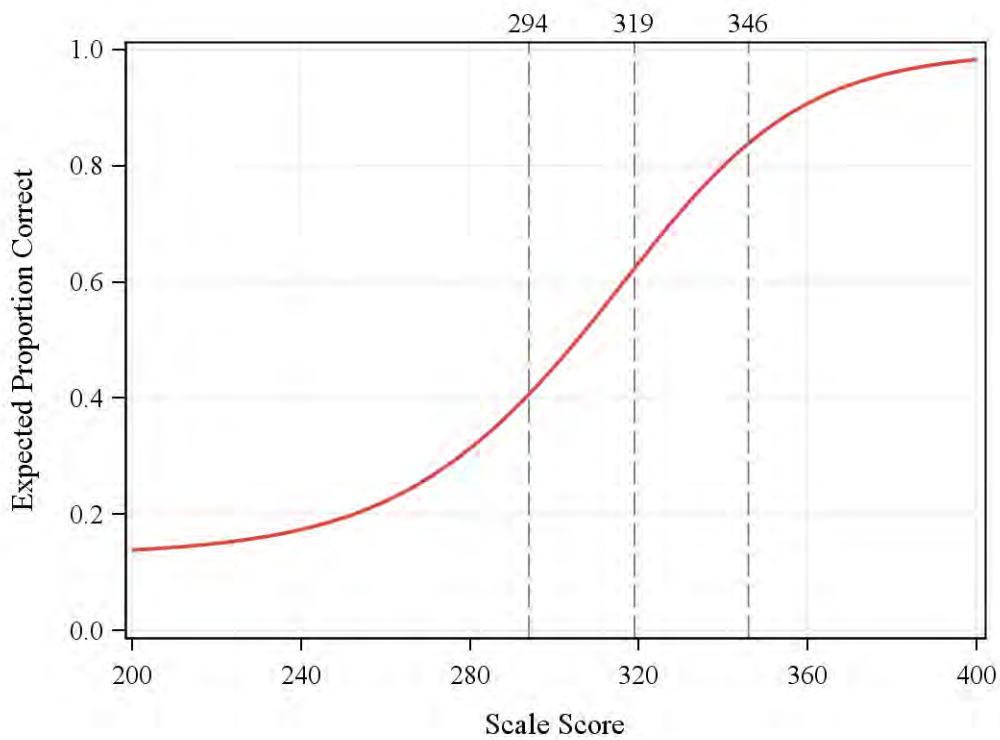


Figure 6.18. Mathematics Grade 5 CSEM Curve

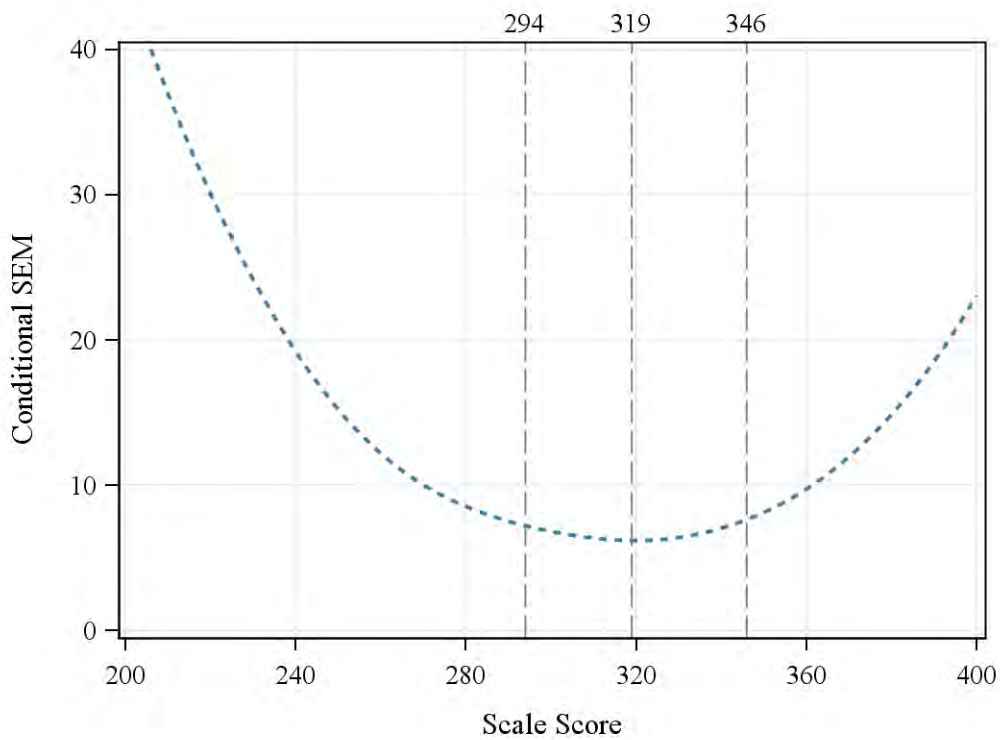


Figure 6.19. Mathematics Grade 6 TCC

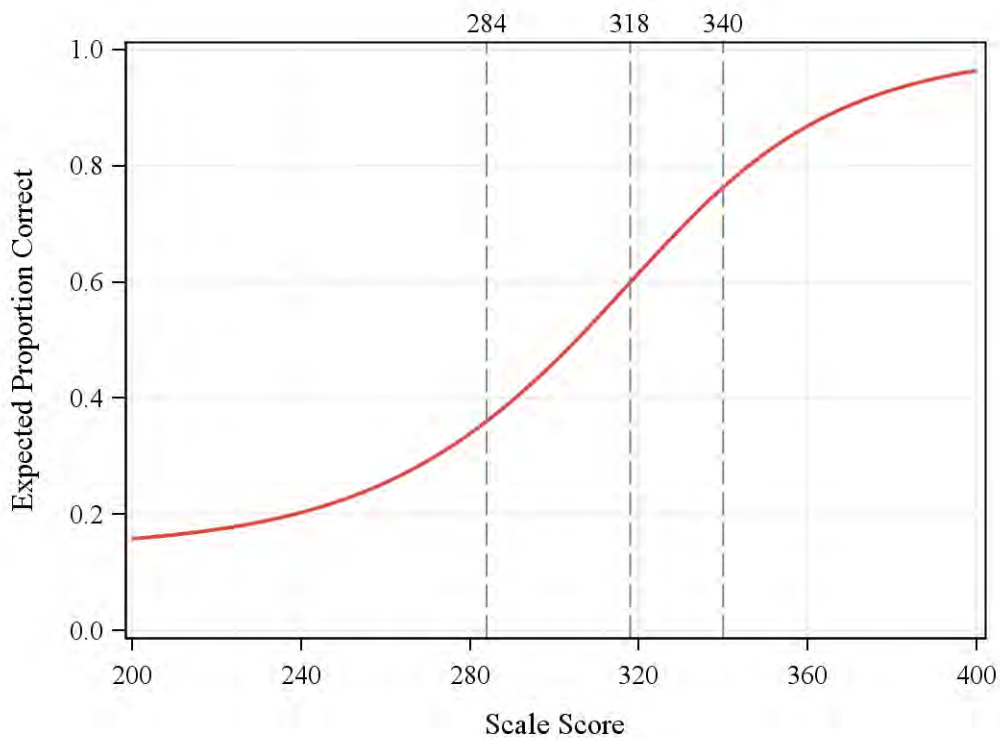


Figure 6.20. Mathematics Grade 6 CSEM Curve

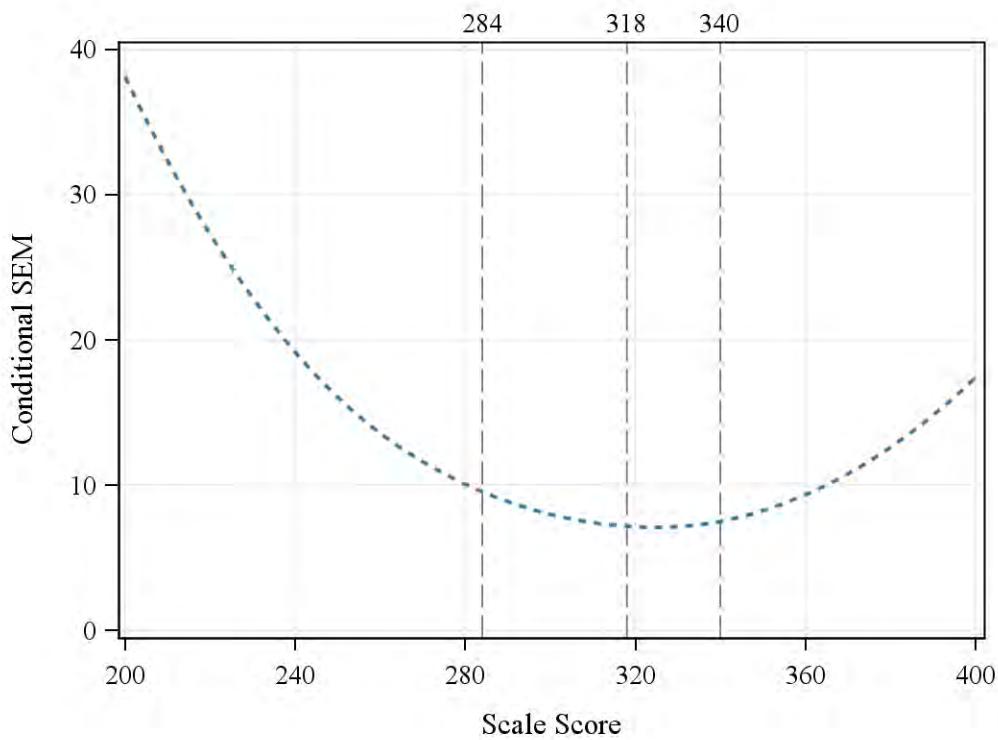


Figure 6.21. Mathematics Grade 7 TCC

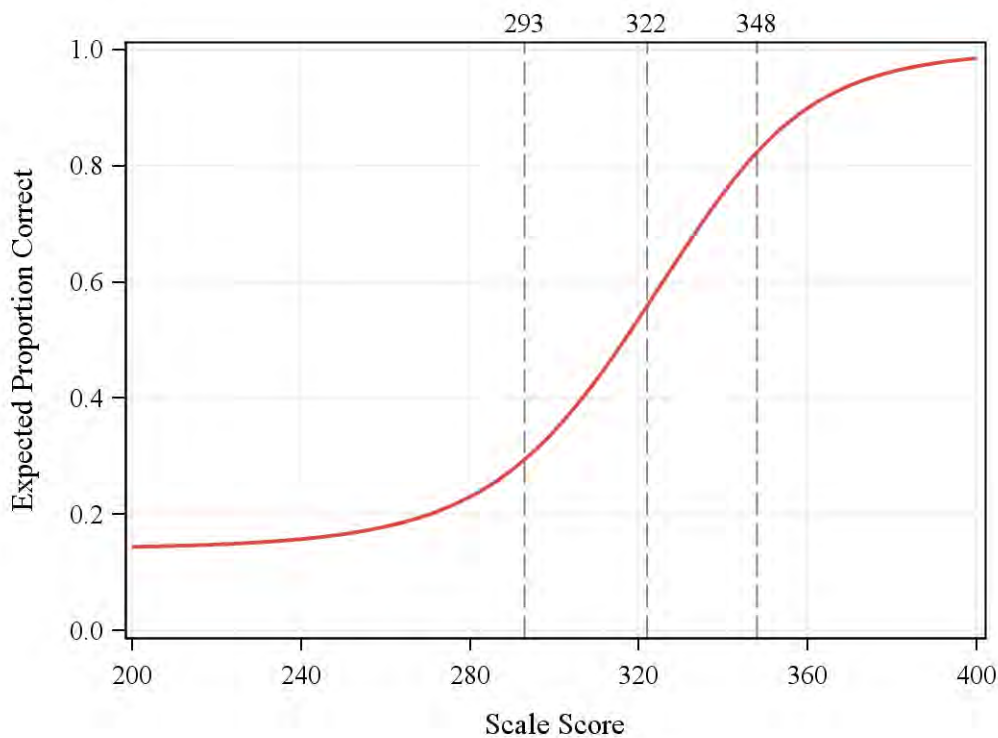


Figure 6.22. Mathematics Grade 7 CSEM Curve

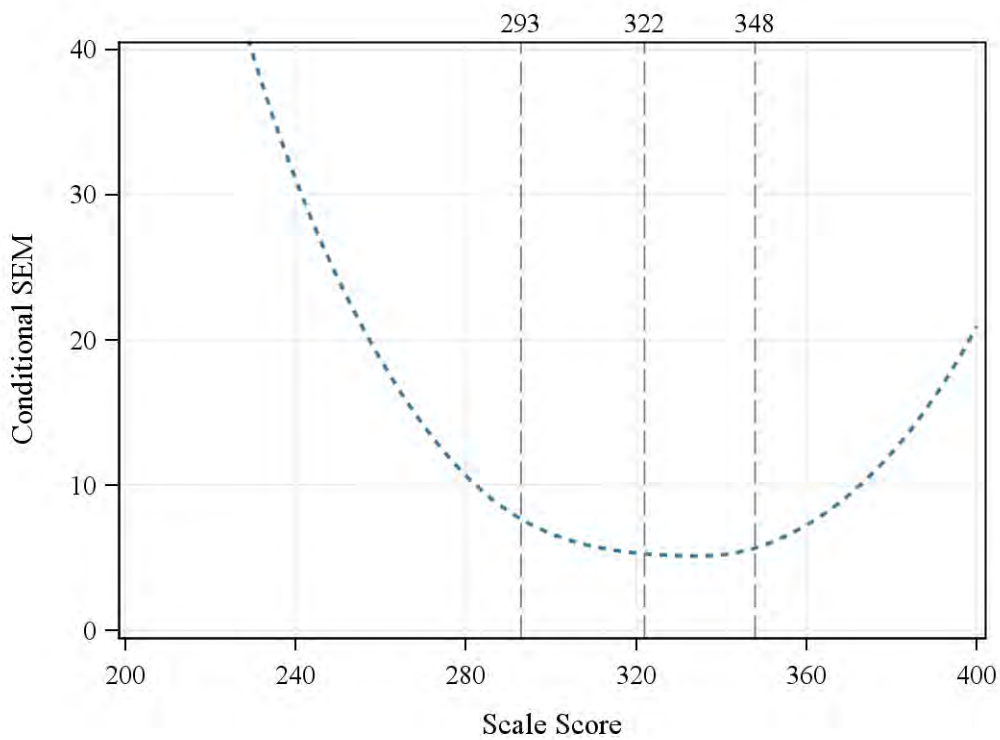


Figure 6.23. Mathematics Grade 8 TCC

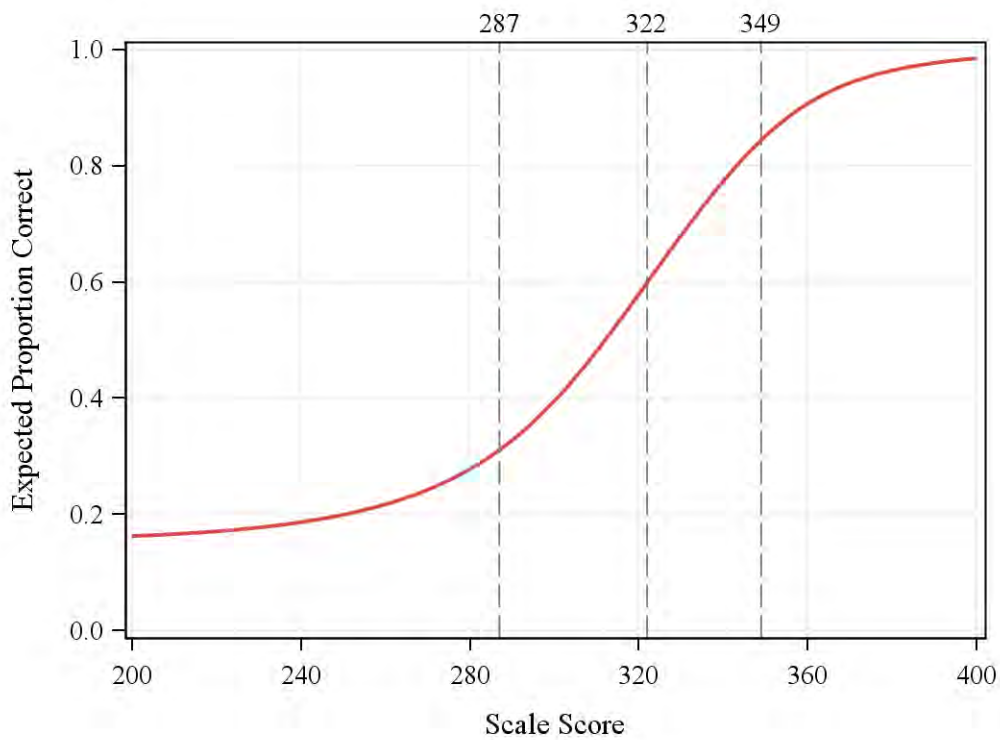
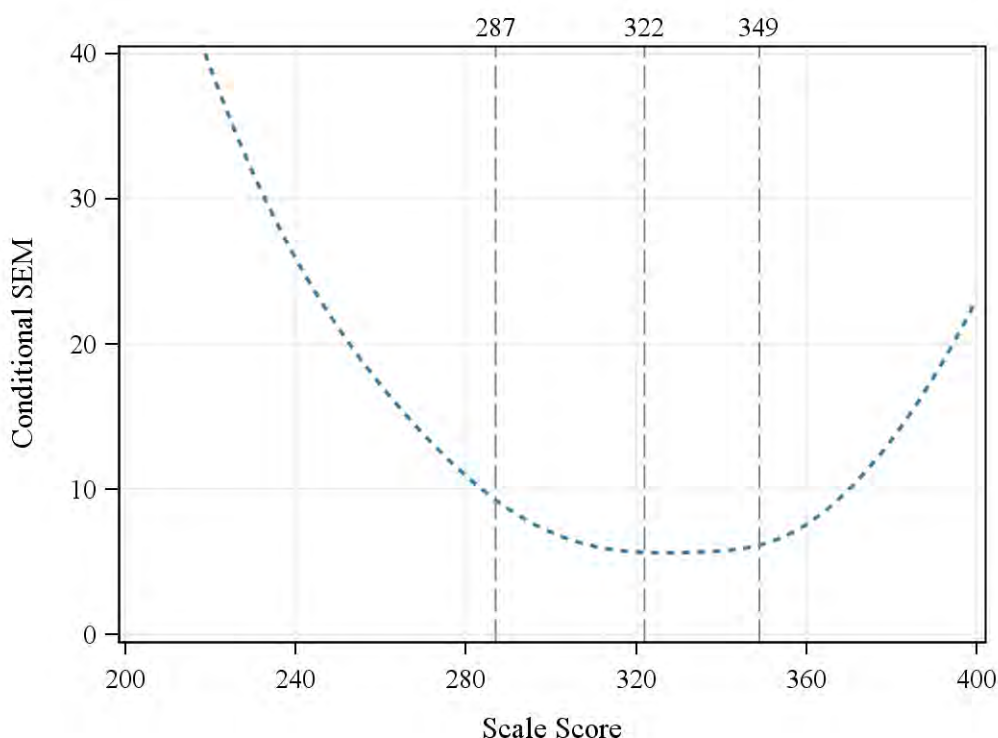


Figure 6.24. Mathematics Grade 8 CSEM Curve



6.8. Scoring Procedure

New York State student examinations were scored using the number correct (NC) scoring method. This method considers how many score points that a student obtained on a test in determining his or her scale score. That is, two students with the same number of score points on the test will receive the same scale score, regardless of which items they answered correctly. In this method, the number correct (or raw) score on the test is converted to a scale score by means of a conversion table. This traditional scoring method is often preferred for its conceptual simplicity and familiarity.

As described in Section 6.5. Linking and Scaling, the final item parameters were used to calculate the raw-score-to-theta tables, using a TCC method (see the details provided below). The obtained scaling transformation intercept and slope (M_1 and M_2) were then applied to the theta values to produce raw score-to-scale score-conversion tables for the Grades 3–8 ELA Tests.

An inverse TCC method was employed using POLYEQUATE (Kolen & Cui, 2004). The inverse of the TCC procedure produces trait values (i.e., proficiency) based on unweighted raw scores. These estimates show negligible statistical bias (defined in statistics as the difference between an estimator's expected value and the true value of the parameter being estimated) for tests with maximum possible raw scores of at least 30 points. All NYSTP ELA and mathematics tests have a maximum raw score higher than 30 points. In the inverse TCC method, a student's trait (i.e., proficiency) estimate is taken to be the trait value that has an expected raw score equal to the student's observed raw score. It was found that, for tests containing only MC items, the inverse of the TCC is an excellent first-order approximation of the number of correct maximum likelihood estimates (MLE) showing negligible bias for tests of at least 30 items. For tests with a mixture of MC and CR items, the MLE and TCC estimates are even more similar (Yen, 1984).

The inverse of the TCC method relies on the following equation:

$$\sum_{i=1}^n v_i x_i = \sum_{i=1}^n v_i E(X_i | \tilde{\theta})$$

where:

x_i is a student's observed raw score on item i ,
 v_i is a non-optimal weight specified in a scoring process ($v_i = 1$ if no weights are specified), and
 $\tilde{\theta}$ is a trait estimate.

Potential differences in test form difficulty at different performance levels are accounted for in the linking and in the resulting raw score-to-scale score conversion tables, so that students of the same proficiency are expected to obtain the same scale score, regardless of which form they took.

6.8.1. Raw Score-to-Scale Score and SEM Conversion Tables

The scale score is the basic score for the NYSTP. Raw score-to-scale score (RSSS) conversion tables based on the total number correct are presented in Appendix Q, Tables Q1–Q12.

The standard error (SE) of a scale score indicates the precision with which the proficiency is estimated, and it inversely is related to the amount of information provided by the test at each performance level. The SE is estimated as follows:

$$SE(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}}$$

where

$SE(\hat{\theta})$ is the standard error of the scale score (theta).

$I(\theta)$ is the amount of information provided by the test at a given performance level.

The information is estimated based on thetas in the scale score metric; therefore, the SE is also expressed in the scale score metric. The SE value varies across performance levels and is the highest at the extreme ends of the scale where the amount of test information is typically the lowest. The final element of the raw score-to-scale score tables is the application of the performance level cut scores.

The linking procedure described above does not guarantee that the same scale score scale points selected as performance-level cut scores will be observed. It was important to appropriately reflect the performance levels set by the standard setting panel and approved by the Commissioner in Summer 2013. To that end, if a given scale score cut was not observed in the 2017 RSSS table, the nearest, but lower, scale score value was rounded up to the established scale score cut. In this way, the approved scale score cuts set in 2013 were maintained for 2017.

Table 6.13 and Table 6.14 present scale score ranges associated with each performance level for ELA and Mathematics, respectively.

Table 6.13. ELA Scale Score Ranges Associated with Each Performance Level

Grade	NYS Level 1	NYS Level 2	NYS Level 3	NYS Level 4
3	180-290	291-319	320-357	358-412
4	164-286	287-319	320-342	343-416
5	126-288	289-319	320-345	346-428
6	128-282	283-319	320-337	338-403
7	133-286	287-317	318-346	347-402
8	121-283	284-315	316-342	343-402

Table 6.14. Mathematics Scale Score Ranges Associated with Each Performance Level

Grade	NYS Level 1	NYS Level 2	NYS Level 3	NYS Level 4
3	145-284	285-313	314-339	340-397
4	133-282	283-313	314-340	341-397
5	151-293	294-318	319-345	346-401
6	132-283	284-317	318-339	340-421
7	160-292	293-321	322-347	348-401
8	134-286	287-321	322-348	349-400

The 2017 administration was the first in which NYSED offered the 3-8 ELA and Mathematics Tests in two administration modes: CBT and PBT. A comparability study was completed to identify whether or not there were any differences in student performance that could be attributed to the mode of test administration (i.e. PBT versus CBT). The main inference to be drawn from the mode comparability study is whether scores that arise from students testing on paper or on computer are interchangeable. A propensity score matching approach was conducted to generate the CBT and PBT samples that were comparable on covariates that may affect student performance, aside from the test mode itself (e.g. gender, school-type, previous performance). The difference in students' test scores were computed between the matched CBT and PBT samples to evaluate test-level mode comparability, and mode adjustments were made accordingly. Please see Appendix R (the mode comparability report) and Appendix S (the NYSED memorandum on the mode comparability results) for more details.

Section 7: Reliability and Standard Error of Measurement

This section presents specific information on various test reliability statistics and standard error of measurement (SEM), as well as the results from a study of performance level classification accuracy and consistency. The data set for these studies includes all tested New York State students who received valid scores.

7.1. Test Reliability

Test reliability is directly related to score stability and standard error and, as such, is an essential element of fairness and validity. Test reliability can be directly measured with an alpha statistic, or the alpha statistic can be used to derive the SEM. For the Grades 3–8 ELA and Mathematics Tests, Questar calculated two types of reliability statistics: Cronbach’s alpha (Cronbach, 1951) and Feldt-Raju coefficient (Qualls, 1995). These two measures are appropriate for assessment of a test’s internal consistency when a single test is administered to a group of examinees on one occasion. The reliability of the test is then estimated by considering how well the items that reflect the same construct yield similar results (or how consistent the results are for different items that reflect the same construct measured by the test). Both Cronbach’s alpha and Feldt-Raju coefficient measures are appropriate for tests of multiple-item formats (MC and CR items).

7.1.1. Test Statistics and Reliability for Total Test

Table 7.1 and Table 7.3 present the test statistics including raw-score (RS) means and raw-score standard deviations (SDs) for ELA and Mathematics, respectively. These statistics give the necessary context for Table 7.2 and Table 7.4, which present the case counts (n-count), number of test items (# Items), Cronbach’s alpha and associated SEM, and Feldt-Raju coefficient and associated SEM obtained for the total ELA and Mathematics tests. Reliability coefficients provide measures of internal consistency that range from zero to one. High reliability indicates that scores are consistent and not unduly influenced by random error. Overall test reliability is a very good indication of each test’s internal consistency.

Grades 3–8 ELA reliability estimates (Cronbach’s alpha and Feldt-Raju) ranged from 0.89 to 0.92. Grades 3–8 Mathematics reliability estimates (Cronbach’s alpha and Feldt-Raju) ranged from 0.93 to 0.95. The reliabilities are similar across grades and slightly higher for the Mathematics tests than for the ELA tests. All reliabilities were at least .89 across all grades and both subjects, which is a good indication that the NYSTP Grades 3–8 ELA and Mathematics Tests are acceptably reliable.

Table 7.1. ELA Test Form Statistics

Grade	Item-Level			Student-Level			
	<i>p</i> -value			N-Count	Raw Score		
	Mean	Min.	Max.		Max.	Mean	SD
3	0.57	0.37	0.84	175,098	47	25.06	9.57
4	0.57	0.33	0.81	174,821	47	25.94	9.05
5	0.59	0.39	0.89	164,596	57	32.61	10.38
6	0.63	0.42	0.84	161,424	57	35.63	11.38
7	0.62	0.41	0.89	152,338	57	36.10	10.88
8	0.67	0.44	0.92	143,207	57	38.76	10.55

Table 7.2. ELA Test Reliability and Standard Error of Measurement

Grade	N-Count	Items	Raw Score	Cronbach's Alpha		Feldt-Raju Coefficient	
			Points	Est.	SEM	Est.	SEM
3	175,098	34	47	0.91	2.89	0.92	2.76
4	174,821	34	47	0.89	3.03	0.90	2.87
5	164,596	44	57	0.89	3.38	0.90	3.23
6	161,424	44	57	0.92	3.25	0.93	3.11
7	152,338	44	57	0.91	3.27	0.92	3.11
8	143,207	44	57	0.91	3.16	0.92	3.01

Table 7.3. Mathematics Test Form Statistics

Grade	Item-Level			Student-Level			
	<i>p</i> -value			N-Count	Raw Score		
	Mean	Min.	Max.		Max.	Mean	SD
3	0.65	0.30	0.94	178,086	56	34.59	13.19
4	0.64	0.32	0.93	176,679	62	37.97	14.82
5	0.60	0.19	0.88	166,570	62	34.03	15.10
6	0.53	0.21	0.85	161,970	68	35.11	15.74
7	0.51	0.22	0.77	142,910	68	32.54	17.44
8	0.46	0.15	0.83	107,670	68	28.74	15.16

Table 7.4. Mathematics Test Reliability and Standard Error of Measurement

Grade	N-Count	Items	Raw Score	Cronbach's Alpha		Feldt-Raju Coefficient	
			Points	Est.	SEM	Est.	SEM
3	178,086	45	56	0.93	3.44	0.94	3.21
4	176,679	48	62	0.94	3.63	0.95	3.33
5	166,570	48	62	0.94	3.63	0.95	3.36

Grade	N-Count	Items	Raw Score	Cronbach's Alpha		Feldt-Raju Coefficient	
			Points	Est.	SEM	Est.	SEM
6	161,970	54	68	0.94	3.81	0.95	3.59
7	142,910	54	68	0.95	3.90	0.96	3.58
8	107,670	54	68	0.94	3.83	0.94	3.63

7.1.2. Reliability of MC Items

In addition to overall test reliability, Cronbach's alpha and Feldt-Raju coefficient were computed separately for MC and CR item sets. It is important to recognize that reliability is directly affected by test length; therefore, reliability estimates for tests by item type will always be lower than reliability estimates for the overall test form. Table 7.5 and Table 7.6 present reliabilities for the subsets of MC items.

Table 7.5. ELA MC Item Reliability and Standard Error of Measurement

Grade	N-Count	Items	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
3	175,098	25	0.84	2.19	0.84	2.18
4	174,821	25	0.81	2.20	0.81	2.19
5	164,596	35	0.82	2.65	0.83	2.64
6	161,424	35	0.88	2.57	0.88	2.56
7	152,338	35	0.86	2.59	0.86	2.57
8	143,207	35	0.86	2.52	0.87	2.50

Table 7.6. Mathematics MC Item Reliability and Standard Error of Measurement

Grade	N-Count	Items	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
3	178,086	37	0.92	2.38	0.92	2.36
4	176,679	38	0.92	2.43	0.92	2.41
5	166,570	38	0.92	2.51	0.92	2.50
6	161,970	44	0.92	2.80	0.92	2.78
7	142,910	44	0.93	2.83	0.93	2.82
8	107,670	44	0.90	2.93	0.90	2.92

7.1.3. Reliability of CR Items

Reliability coefficients were also computed for the subsets of CR items. The results are presented in Table 7.7 and Table 7.8.

Table 7.7. ELA CR Item Reliability and Standard Error of Measurement

Grade	N-Count	Items	Raw Score Points	Cronbach's Alpha		Feldt-Raju Coefficient	
				Est.	SEM	Est.	SEM
3	175,098	9	22	0.88	1.65	0.89	1.58
4	174,821	9	22	0.85	1.86	0.86	1.77
5	164,596	9	22	0.86	1.82	0.87	1.75
6	161,424	9	22	0.89	1.68	0.90	1.60
7	152,338	9	22	0.89	1.68	0.90	1.57
8	143,207	9	22	0.87	1.65	0.89	1.53

Note. Results should be interpreted with caution because the number of items is low.

Table 7.8. Mathematics CR Item Reliability and Standard Error of Measurement

Grade	N-Count	Items	Raw Score Points	Cronbach's Alpha		Feldt-Raju Coefficient	
				Est.	SEM	Est.	SEM
3	178,086	8	19	0.84	2.22	0.84	2.18
4	176,679	10	24	0.89	2.35	0.89	2.27
5	166,570	10	24	0.90	2.26	0.90	2.19
6	161,970	10	24	0.88	2.32	0.89	2.24
7	142,910	10	24	0.91	2.24	0.92	2.16
8	107,670	10	24	0.89	2.14	0.89	2.12

Note. Results should be interpreted with caution because the number of items is low.

7.1.4. Test Reliability for Reporting Categories

In this section, reliability coefficients that were estimated for the population and subgroups are presented. The reporting categories include the following: gender, ethnicity, NRC, ELL, all SWD, all SUA, SWD/SUA (includes examinees who are classified as having a disability and who use at least one disability-related accommodation), and English language learners using accommodations specific to their ELL status (ELL/SUA). Accommodations available to students include the following: Flexibility in Scheduling/Timing, Flexibility in Setting, Method of Presentation (excluding Braille), Method of Response, Braille and Large-type, and others. Accommodations available to English language learners are Separate Location, and Bilingual Dictionaries and Glossaries.

As shown in Tables 7.9–7.14 and Tables 7.15–7.20 for ELA and Mathematics, respectively, the estimated reliabilities for subgroups were close in magnitude to the test reliability estimates of the population. Cronbach's alpha reliability coefficients were all at least 0.78. Feldt-Raju reliability coefficients, which tend to be larger than the Cronbach's alpha estimates for the same group, were at least 0.79 each. These indicate a very good test internal consistency (reliability) for analyzed subgroups of examinees.

Table 7.9. ELA Grade 3 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	175,098	0.91	2.89	0.92	2.76
Gender	Female	86,944	0.91	2.88	0.92	2.75
	Male	88,154	0.91	2.89	0.92	2.76
Ethnicity	Asian	17,794	0.90	2.85	0.91	2.70
	Black	31,280	0.91	2.93	0.92	2.79
	Hispanic	48,492	0.90	2.91	0.91	2.79
	American Indian	1,152	0.90	2.94	0.91	2.80
	Multiracial	5,122	0.92	2.86	0.92	2.71
	Pacific Islander	488	0.90	2.90	0.91	2.78
	White	70,770	0.90	2.85	0.91	2.72
NRC	New York	64,712	0.91	2.90	0.92	2.75
	Big 4 Cities	7,726	0.91	2.90	0.91	2.78
	Urban/Suburban	14,236	0.90	2.88	0.91	2.79
	Rural	9,868	0.90	2.88	0.90	2.78
	Average Needs	40,722	0.90	2.85	0.91	2.74
	Low Needs	17,939	0.88	2.73	0.89	2.63
	Charter School	11,164	0.89	2.90	0.89	2.79
	Religious and Independent	8,731	0.91	2.96	0.92	2.79
SWD	All Codes	22,349	0.88	2.85	0.89	2.76
SUA	All Codes	19,269	0.88	2.84	0.88	2.76
ELL	ELL=Y	18,531	0.86	2.91	0.87	2.81
SWD/SUA	SWD & SUA codes	16,952	0.87	2.83	0.88	2.76
ELL/SUA	SUA & ELL codes	2,972	0.84	2.80	0.84	2.74

Table 7.10. ELA Grade 4 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	174,821	0.89	3.03	0.90	2.87
Gender	Female	87,697	0.89	3.01	0.90	2.86
	Male	87,124	0.89	3.02	0.90	2.88
Ethnicity	Asian	18,399	0.89	2.92	0.90	2.76
	Black	31,452	0.88	3.06	0.89	2.90
	Hispanic	48,380	0.88	3.04	0.89	2.91
	American Indian	1,170	0.88	3.05	0.90	2.89
	Multiracial	4,542	0.89	3.01	0.91	2.84
	Pacific Islander	548	0.87	2.97	0.88	2.86

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
Ethnicity	White	70,330	0.88	3.00	0.90	2.85
NRC	New York	65,999	0.89	3.00	0.90	2.84
	Big 4 Cities	7,537	0.88	3.06	0.90	2.90
	Urban/Suburban	13,332	0.87	3.05	0.89	2.92
	Rural	9,548	0.87	3.04	0.89	2.91
	Average Needs	38,875	0.87	3.01	0.89	2.88
	Low Needs	17,686	0.86	2.89	0.87	2.78
	Charter School	9,500	0.86	3.01	0.87	2.90
	Religious and Independent	12,344	0.88	3.09	0.90	2.90
SWD	All Codes	23,099	0.86	2.98	0.87	2.87
SUA	All Codes	21,442	0.85	2.98	0.86	2.87
ELL	ELL=Y	14,963	0.82	3.02	0.83	2.90
SWD/SUA	SWD & SUA codes	18,619	0.85	2.96	0.86	2.86
ELL/SUA	SUA & ELL codes	3,086	0.78	2.92	0.79	2.83

Table 7.11. ELA Grade 5 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	164,596	0.89	3.38	0.90	3.23
Gender	Female	81,838	0.89	3.35	0.90	3.21
	Male	82,758	0.89	3.38	0.90	3.24
Ethnicity	Asian	17,365	0.89	3.25	0.90	3.11
	Black	30,504	0.89	3.43	0.90	3.28
	Hispanic	45,194	0.88	3.41	0.89	3.28
	American Indian	1,123	0.89	3.43	0.90	3.26
	Multiracial	3,882	0.90	3.35	0.91	3.19
	Pacific Islander	606	0.88	3.35	0.89	3.23
	White	65,922	0.89	3.34	0.90	3.19
NRC	New York	63,287	0.90	3.36	0.91	3.21
	Big 4 Cities	6,915	0.89	3.43	0.90	3.28
	Urban/Suburban	12,618	0.88	3.41	0.89	3.29
	Rural	9,115	0.88	3.38	0.89	3.26
	Average Needs	37,090	0.88	3.33	0.89	3.21
	Low Needs	17,300	0.86	3.20	0.87	3.10
	Charter School	9,815	0.88	3.39	0.88	3.28
	Religious and Independent	8,456	0.90	3.46	0.91	3.26
SWD	All Codes	23,265	0.86	3.39	0.87	3.28

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
SUA	All Codes	22,111	0.86	3.39	0.87	3.29
ELL	ELL=Y	12,894	0.81	3.41	0.82	3.30
SWD/SUA	SWD & SUA codes	19,028	0.85	3.38	0.86	3.28
ELL/SUA	SUA & ELL codes	2,940	0.78	3.32	0.79	3.25

Table 7.12. ELA Grade 6 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	161,424	0.92	3.25	0.93	3.11
Gender	Female	79,546	0.91	3.19	0.92	3.07
	Male	81,878	0.92	3.27	0.93	3.13
Ethnicity	Asian	17,207	0.92	3.02	0.92	2.89
	Black	31,028	0.91	3.33	0.92	3.19
	Hispanic	44,338	0.91	3.31	0.92	3.18
	American Indian	1,138	0.92	3.30	0.93	3.15
	Multiracial	3,280	0.92	3.21	0.93	3.05
	Pacific Islander	478	0.91	3.20	0.92	3.05
	White	63,955	0.91	3.20	0.92	3.05
NRC	New York	62,289	0.92	3.22	0.93	3.07
	Big 4 Cities	6,534	0.92	3.38	0.93	3.22
	Urban/Suburban	11,343	0.91	3.37	0.92	3.23
	Rural	8,250	0.91	3.32	0.92	3.17
	Average Needs	34,490	0.91	3.22	0.92	3.08
	Low Needs	16,496	0.89	3.04	0.90	2.93
	Charter School	10,412	0.90	3.22	0.91	3.14
	Religious and Independent	11,610	0.91	3.32	0.92	3.15
SWD	All Codes	23,727	0.88	3.37	0.89	3.26
SUA	All Codes	21,872	0.89	3.37	0.89	3.26
ELL	ELL=Y	11,553	0.85	3.39	0.87	3.25
SWD/SUA	SWD & SUA codes	18,939	0.88	3.37	0.89	3.26
ELL/SUA	SUA & ELL codes	2,591	0.81	3.31	0.82	3.21

Table 7.13. ELA Grade 7 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	152,338	0.91	3.27	0.92	3.11
Gender	Female	74,881	0.90	3.18	0.91	3.06

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
Gender	Male	77,457	0.91	3.30	0.92	3.13
Ethnicity	Asian	17,121	0.91	3.00	0.92	2.87
	Black	29,097	0.89	3.35	0.90	3.20
	Hispanic	40,403	0.90	3.32	0.91	3.18
	American Indian	1,110	0.90	3.31	0.90	3.17
	Multiracial	2,845	0.92	3.26	0.93	3.07
	Pacific Islander	435	0.90	3.19	0.91	3.04
	White	61,327	0.91	3.23	0.92	3.06
NRC	New York	61,134	0.91	3.19	0.92	3.05
	Big 4 Cities	6,006	0.90	3.43	0.91	3.25
	Urban/Suburban	10,238	0.90	3.39	0.91	3.23
	Rural	7,926	0.90	3.34	0.91	3.19
	Average Needs	31,261	0.91	3.26	0.91	3.11
	Low Needs	17,065	0.89	3.06	0.89	2.95
	Charter School	9,977	0.88	3.24	0.88	3.16
	Religious and Independent	8,731	0.91	3.39	0.93	3.16
SWD	All Codes	22,363	0.87	3.40	0.88	3.26
SUA	All Codes	20,616	0.88	3.40	0.89	3.26
ELL	ELL=Y	10,895	0.84	3.44	0.86	3.25
SWD/SUA	SWD & SUA codes	17,739	0.86	3.41	0.87	3.27
ELL/SUA	SUA & ELL codes	2,463	0.78	3.36	0.80	3.23

Table 7.14. ELA Grade 8 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	143,207	0.91	3.16	0.92	3.01
Gender	Female	69,786	0.90	3.05	0.91	2.93
	Male	73,421	0.91	3.22	0.92	3.07
Ethnicity	Asian	16,568	0.91	2.88	0.92	2.75
	Black	29,194	0.90	3.25	0.91	3.11
	Hispanic	39,307	0.90	3.23	0.91	3.09
	American Indian	1,139	0.91	3.22	0.92	3.06
	Multiracial	2,129	0.92	3.19	0.93	2.98
	Pacific Islander	435	0.91	3.08	0.92	2.93
	White	54,435	0.91	3.11	0.92	2.96
NRC	New York	60,999	0.91	3.11	0.92	2.97
	Big 4 Cities	5,963	0.91	3.44	0.92	3.23

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
NRC	Urban/Suburban	9,352	0.91	3.32	0.92	3.17
	Rural	7,355	0.91	3.27	0.91	3.12
	Average Needs	27,774	0.91	3.18	0.92	3.02
	Low Needs	14,394	0.89	2.95	0.90	2.83
	Charter School	8,191	0.87	3.04	0.88	2.99
	Religious and Independent	9,179	0.89	3.10	0.90	2.98
SWD	All Codes	20,479	0.88	3.40	0.89	3.27
SUA	All Codes	19,038	0.89	3.39	0.90	3.26
ELL	ELL=Y	9,433	0.85	3.46	0.86	3.29
SWD/SUA	SWD & SUA codes	16,357	0.88	3.41	0.89	3.27
ELL/SUA	SUA & ELL codes	2,061	0.83	3.43	0.84	3.29

Table 7.15. Mathematics Grade 3 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	178,086	0.93	3.44	0.94	3.21
Gender	Female	87,363	0.93	3.44	0.94	3.22
	Male	90,723	0.93	3.43	0.94	3.21
Ethnicity	Asian	18,619	0.92	3.18	0.94	2.93
	Black	31,919	0.93	3.46	0.94	3.25
	Hispanic	50,809	0.93	3.47	0.93	3.28
	American Indian	1,188	0.93	3.46	0.94	3.25
	Multiracial	4,784	0.94	3.43	0.94	3.19
	Pacific Islander	511	0.93	3.42	0.94	3.19
	White	70,256	0.92	3.42	0.93	3.20
NRC	New York	68,518	0.93	3.43	0.94	3.20
	Big 4 Cities	7,783	0.93	3.45	0.94	3.26
	Urban/Suburban	14,259	0.93	3.47	0.94	3.28
	Rural	9,702	0.93	3.49	0.93	3.29
	Average Needs	40,506	0.92	3.45	0.93	3.25
	Low Needs	17,950	0.91	3.28	0.92	3.09
	Charter School	11,587	0.93	3.25	0.94	3.01
	Religious and Independent	7,781	0.93	3.52	0.93	3.31
SWD	All Codes	24,609	0.93	3.41	0.93	3.26
SUA	All Codes	22,553	0.92	3.41	0.93	3.27
ELL	ELL=Y	21,146	0.92	3.45	0.93	3.29
SWD/SUA	SWD & SUA codes	20,038	0.92	3.39	0.93	3.26

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
ELL/SUA	SUA & ELL codes	3,736	0.91	3.35	0.92	3.23
ELL Test Language	English	173,708	0.93	3.44	0.94	3.21
	Chinese	619	0.92	3.28	0.93	3.05
	Haitian-Creole	68	0.88	3.21	0.89	3.08
	Korean	42	0.92	3.44	0.93	3.18
	Russian	125	0.92	3.50	0.93	3.32
	Spanish	3,524	0.91	3.39	0.91	3.26
	All Translations	4,378	0.93	3.42	0.94	3.25

Table 7.16. Mathematics Grade 4 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	176,679	0.94	3.63	0.95	3.33
Gender	Female	87,324	0.94	3.62	0.95	3.34
	Male	89,355	0.94	3.63	0.95	3.33
Ethnicity	Asian	19,258	0.94	3.31	0.95	3.01
	Black	31,928	0.94	3.66	0.95	3.41
	Hispanic	50,248	0.93	3.67	0.94	3.43
	American Indian	1,190	0.94	3.65	0.95	3.38
	Multiracial	4,301	0.94	3.59	0.95	3.29
	Pacific Islander	569	0.93	3.53	0.94	3.25
	White	69,185	0.93	3.56	0.94	3.30
NRC	New York	69,786	0.94	3.62	0.95	3.32
	Big 4 Cities	7,534	0.94	3.60	0.95	3.37
	Urban/Suburban	13,256	0.94	3.66	0.94	3.42
	Rural	9,323	0.93	3.67	0.94	3.41
	Average Needs	38,564	0.93	3.61	0.94	3.35
	Low Needs	17,838	0.92	3.39	0.93	3.16
	Charter School	9,659	0.94	3.49	0.95	3.20
	Religious and Independent	10,719	0.93	3.73	0.94	3.47
SWD	All Codes	25,094	0.93	3.60	0.94	3.40
SUA	All Codes	24,404	0.93	3.61	0.94	3.41
ELL	ELL=Y	17,318	0.93	3.58	0.93	3.40
SWD/SUA	SWD & SUA codes	21,316	0.92	3.58	0.93	3.40
ELL/SUA	SUA & ELL codes	3,785	0.91	3.48	0.91	3.36
ELL Test Language	English	172,857	0.94	3.62	0.95	3.33
	Chinese	618	0.92	3.49	0.93	3.21

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
ELL Test Language	Haitian-Creole	78	0.91	3.24	0.92	3.14
	Korean	35	0.95	3.36	0.96	3.00
	Russian	110	0.94	3.65	0.95	3.35
	Spanish	2,981	0.92	3.50	0.92	3.36
	All Translations	3,822	0.94	3.58	0.95	3.35

Table 7.17. Mathematics Grade 5 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	166,570	0.94	3.63	0.95	3.36
Gender	Female	82,030	0.94	3.64	0.95	3.37
	Male	84,540	0.94	3.62	0.95	3.34
Ethnicity	Asian	18,043	0.94	3.45	0.95	3.12
	Black	30,690	0.93	3.57	0.94	3.38
	Hispanic	46,789	0.93	3.61	0.94	3.40
	American Indian	1,141	0.94	3.61	0.95	3.36
	Multiracial	3,619	0.95	3.62	0.95	3.33
	Pacific Islander	623	0.94	3.64	0.95	3.33
	White	65,665	0.94	3.62	0.94	3.37
NRC	New York	66,795	0.94	3.63	0.95	3.33
	Big 4 Cities	6,940	0.94	3.44	0.94	3.27
	Urban/Suburban	12,425	0.93	3.56	0.94	3.37
	Rural	8,770	0.93	3.60	0.94	3.41
	Average Needs	36,368	0.93	3.63	0.94	3.39
	Low Needs	17,260	0.92	3.53	0.93	3.28
	Charter School	9,860	0.93	3.61	0.94	3.35
	Religious and Independent	8,152	0.93	3.69	0.94	3.47
SWD	All Codes	24,926	0.92	3.40	0.93	3.27
SUA	All Codes	24,891	0.93	3.41	0.93	3.29
ELL	ELL=Y	15,224	0.92	3.43	0.92	3.30
SWD/SUA	SWD & SUA codes	21,705	0.92	3.37	0.92	3.26
ELL/SUA	SUA & ELL codes	3,576	0.90	3.26	0.90	3.19
ELL Test Language	English	162,853	0.94	3.63	0.95	3.36
	Chinese	655	0.93	3.57	0.94	3.32
	Haitian-Creole	81	0.88	3.10	0.88	3.02
	Korean	17	0.90	3.44	0.92	3.09
	Russian	104	0.93	3.65	0.94	3.42

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
ELL Test Language	Spanish	2,860	0.89	3.30	0.90	3.21
	All Translations	3,717	0.94	3.47	0.94	3.27

Table 7.18. Mathematics Grade 6 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	161,970	0.94	3.81	0.95	3.59
Gender	Female	79,222	0.94	3.81	0.95	3.59
	Male	82,748	0.94	3.81	0.95	3.58
Ethnicity	Asian	17,755	0.95	3.56	0.95	3.32
	Black	30,706	0.93	3.80	0.93	3.61
	Hispanic	44,907	0.93	3.80	0.93	3.64
	American Indian	1,136	0.94	3.85	0.94	3.62
	Multiracial	3,044	0.94	3.79	0.95	3.54
	Pacific Islander	492	0.94	3.81	0.95	3.55
	White	63,930	0.94	3.80	0.94	3.58
NRC	New York	64,195	0.95	3.80	0.95	3.55
	Big 4 Cities	6,498	0.93	3.74	0.94	3.54
	Urban/Suburban	10,914	0.92	3.75	0.93	3.61
	Rural	7,902	0.93	3.84	0.93	3.66
	Average Needs	33,550	0.93	3.81	0.94	3.61
	Low Needs	16,309	0.93	3.64	0.93	3.46
	Charter School	10,518	0.94	3.79	0.94	3.59
	Religious and Independent	12,084	0.93	3.91	0.94	3.68
SWD	All Codes	23,805	0.90	3.67	0.91	3.54
SUA	All Codes	22,959	0.90	3.70	0.91	3.55
ELL	ELL=Y	13,820	0.91	3.62	0.91	3.55
SWD/SUA	SWD & SUA codes	19,789	0.89	3.66	0.90	3.53
ELL/SUA	SUA & ELL codes	2,929	0.84	3.47	0.85	3.44
ELL Test Language	English	157,394	0.94	3.82	0.95	3.59
	Chinese	740	0.94	3.80	0.94	3.55
	Haitian-Creole	83	0.83	3.43	0.84	3.37
	Korean	19	0.96	3.46	0.97	3.16
	Russian	113	0.93	3.70	0.94	3.53
	Spanish	3,621	0.86	3.41	0.87	3.33
	All Translations	4,576	0.93	3.44	0.93	3.47

Table 7.19. Mathematics Grade 7 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	142,910	0.95	3.90	0.96	3.58
Gender	Female	69,674	0.95	3.90	0.96	3.59
	Male	73,236	0.95	3.89	0.96	3.57
Ethnicity	Asian	17,079	0.95	3.71	0.96	3.37
	Black	27,800	0.94	3.74	0.94	3.53
	Hispanic	39,562	0.94	3.81	0.94	3.58
	American Indian	1,100	0.94	3.87	0.95	3.61
	Multiracial	2,251	0.95	3.89	0.96	3.54
	Pacific Islander	429	0.95	3.92	0.95	3.62
	White	54,689	0.94	3.94	0.95	3.65
NRC	New York	62,622	0.95	3.86	0.96	3.52
	Big 4 Cities	6,003	0.93	3.56	0.93	3.39
	Urban/Suburban	8,551	0.92	3.72	0.93	3.53
	Rural	7,066	0.93	3.91	0.94	3.67
	Average Needs	26,242	0.94	3.96	0.95	3.68
	Low Needs	15,989	0.94	3.86	0.95	3.59
	Charter School	9,930	0.95	3.86	0.95	3.60
	Religious and Independent	6,507	0.94	3.97	0.94	3.71
SWD	All Codes	21,507	0.91	3.47	0.91	3.36
SUA	All Codes	20,559	0.92	3.52	0.92	3.39
ELL	ELL=Y	12,074	0.92	3.49	0.92	3.36
SWD/SUA	SWD & SUA codes	17,925	0.90	3.45	0.91	3.34
ELL/SUA	SUA & ELL codes	2,617	0.82	3.23	0.82	3.19
ELL Test Language	English	138,544	0.95	3.90	0.96	3.59
	Chinese	777	0.95	3.82	0.95	3.51
	Haitian-Creole	86	0.90	3.40	0.91	3.28
	Korean	27	0.95	3.85	0.96	3.48
	Russian	126	0.93	3.95	0.94	3.66
	Spanish	3,350	0.86	3.36	0.87	3.30
	All Translations	4,366	0.94	3.59	0.95	3.36

Table 7.20. Mathematics Grade 8 Test Reliability by Subgroup

Demographic Category		N-Count	Cronbach's Alpha		Feldt-Raju Coefficient	
			Est.	SEM	Est.	SEM
State	All Items	107,670	0.94	3.83	0.94	3.63
Gender	Female	51,545	0.94	3.86	0.94	3.65
	Male	56,125	0.94	3.78	0.94	3.61
Ethnicity	Asian	11,110	0.95	3.84	0.96	3.52
	Black	23,673	0.92	3.68	0.93	3.55
	Hispanic	33,077	0.93	3.74	0.93	3.60
	American Indian	901	0.93	3.73	0.93	3.57
	Multiracial	1,374	0.94	3.85	0.95	3.62
	Pacific Islander	345	0.94	3.84	0.95	3.61
	White	37,190	0.93	3.91	0.93	3.73
NRC	New York	51,257	0.94	3.81	0.95	3.58
	Big 4 Cities	5,302	0.91	3.50	0.91	3.41
	Urban/Suburban	6,359	0.88	3.56	0.89	3.50
	Rural	5,411	0.90	3.79	0.91	3.67
	Average Needs	16,047	0.91	3.87	0.91	3.73
	Low Needs	7,562	0.92	3.97	0.93	3.79
	Charter School	6,398	0.94	3.78	0.95	3.58
	Religious and Independent	9,334	0.93	3.93	0.94	3.74
SWD	All Codes	18,050	0.87	3.45	0.88	3.40
SUA	All Codes	17,463	0.89	3.48	0.89	3.42
ELL	ELL=Y	10,765	0.91	3.52	0.91	3.46
SWD/SUA	SWD & SUA codes	15,199	0.87	3.44	0.87	3.39
ELL/SUA	SUA & ELL codes	2,034	0.80	3.27	0.80	3.27
ELL Test Language	English	103,722	0.94	3.84	0.94	3.64
	Chinese	658	0.95	3.92	0.96	3.57
	Haitian-Creole	75	0.79	3.43	0.80	3.33
	Korean	9	0.95	3.50	0.96	2.96
	Russian	114	0.93	3.85	0.94	3.65
	Spanish	3,092	0.84	3.36	0.85	3.34
	All Translations	3,948	0.94	3.51	0.94	3.42

7.2. Standard Error of Measurement (SEM)

Table 7.2 and Table 7.4 present the SEMs, as computed from Cronbach's alpha and the Feldt-Raju reliability statistics, for ELA and Mathematics, respectively. The SEMs ranged from 2.76 to 3.90 across subjects, grades, and the two methods of estimation, which is reasonable and small. The SEMs are directly related to reliability: the higher the reliability, the lower the standard

error. As discussed, the reliability of these tests is relatively high, so it was expected that the SEMs would be very low.

The SEMs for subpopulations, as computed from Cronbach's alpha and the Feldt-Raju reliability statistics, are presented in Tables 7.9–7.14 and Tables 7.15–7.20. The SEMs associated with all reliability estimates for all subjects, grades, methods of estimation, and subpopulations ranged from 2.63 to 3.97, which is acceptably close to those for the entire population. This narrow range indicates that across the Grades 3–8 ELA and Mathematics Tests, all students' test scores are reasonably reliable with minimal error.

7.3. Performance Level Classification Consistency and Accuracy

This subsection describes the analyses conducted to estimate performance level classification consistency and accuracy for the Grades 3–8 ELA and Mathematics Tests. In other words, this provides statistical information on the classification of students into the four performance categories. Classification consistency refers to the estimated degree of agreement between examinees' performance classification from two independent administrations of the same test (or from two parallel forms of the test). Because obtaining test scores from two independent administrations of New York State tests was not feasible due to item release after each administration, a psychometric model was used to obtain the estimated classification consistency indices, using test scores from a single administration. Classification accuracy can be defined as the agreement between the actual classifications using observed cut scores and true classifications based on known true cut scores (Livingston and Lewis, 1995).

In conjunction with measures of internal consistency, classification consistency is an important type of reliability and is particularly relevant to high-stakes pass/fail tests. As a form of reliability, classification consistency represents how reliably students can be classified into performance categories.

Classification consistency is most relevant for students whose proficiency is near the pass/fail cut score. For example, consider the cut score delineating Levels II and III or simply the "Level III Cut." Students whose proficiency is far above or far below that cut score are unlikely to be misclassified because repeated administration of the test will nearly always result in the same classification. Examinees whose true scores are close to the cut score are a more serious concern. These students' true scores will likely lie within the SEM of the cut score. For this reason, the measurement error at the cut scores should be considered when evaluating the classification consistency of a test. Furthermore, the number of students near the cut scores should also be considered when evaluating classification consistency; these numbers show the number of students who are at risk of being misclassified. Scoring tables with SEMs are located in Section 6: IRT Calibration and Scaling, and student scale score frequency distributions are located in Appendix Q. Classification consistency and accuracy were estimated using the IRT procedure suggested by Lee, Hanson, and Brennan (2002) and Wang, Kolen, and Harris (2000). Appendix P includes a description of the calculations and procedure based on the paper by Lee et al. (2002).

7.3.1. Consistency

The results for classifying students into four performance levels are separated from results based solely on the Level III cut. Table 7.21 and Table 7.22 include case counts (n-count),

classification consistency (Agreement), classification inconsistency (Inconsistency), and Cohen's kappa (Kappa). Consistency indicates the rate at which a second administration would yield the same performance category designation (or a different designation for the inconsistency rate). The agreement index is a sum of the diagonal element in the contingency table. Kappa is similar, but corrects for chance agreement. The inconsistency index is equal to the "1 - agreement index."

Table 7.21 depicts the ELA and Mathematics consistency study results, based on the range of performance levels for all grades. For ELA, 69–75% of students were estimated to be classified consistently to one of the four performance categories with a hypothetical second administration. Kappa—which corrects for chance agreement—ranged from 0.56 to 0.63. These are between “moderate” and “substantial” agreement, as per Landis and Koch's (1977) rules of thumb for kappa. For Mathematics, 76–81% of students were estimated to be classified consistently to one of the four performance categories, and kappa ranged from 0.68 to 0.72. These are all considered “substantial” agreement, by Landis and Koch's (1977) rules of thumb for the kappa statistic. As mentioned above and for all tests, there is an acceptable amount of measurement error that all scores contain. By random chance, students testing twice may be classified first, for example, as a Level III and second as a Level IV. This is expected to occur more often for students scoring around the selected cut score, and less often for students closer to the middle of the performance level (i.e., close to the mid-point of two adjacent cut scores).

Table 7.21. Decision Consistency (All Cuts)*

Grade	N-Count	Agreement	Inconsistency	Kappa
ELA				
3	175,098	75%	25%	0.62
4	174,819	72%	28%	0.59
5	164,594	69%	31%	0.56
6	161,424	73%	27%	0.61
7	152,338	74%	26%	0.63
8	143,206	72%	28%	0.61
Math				
3	178,085	76%	24%	0.67
4	176,677	78%	22%	0.69
5	166,570	81%	19%	0.72
6	161,968	77%	23%	0.68
7	142,910	81%	19%	0.72
8	107,669	79%	21%	0.70

*Note: Decision consistency was calculated for PBT students only as item parameters were disproportionately based on PBT.

Table 7.22 depicts the ELA and Mathematics consistency study results based on two performance levels (NYS Level II and NYS Level III) as defined by the Level III cut. For ELA, 92–99% of the classifications of individual students were estimated to remain stable with a second administration. Kappa coefficients for ELA classification consistency ranged from 0.60 to 0.71. These are considered “substantial” agreement, as per Landis and Koch's (1977) rules of

thumb for kappa. For Mathematics, 93–97% of the classifications were estimated consistently, and kappa coefficients ranged from 0.77 to 0.82. As with ELA, these statistics indicate at least “substantial” agreement (where kappa > 0.60) and some indicating “almost perfect” agreement (where kappa > 0.80), as per Landis and Koch’s (1977) rules of thumb for kappa.

Table 7.22. Decision Consistency (Level III Cut)*

Grade	N-Count	Agreement	Inconsistency	Kappa
ELA				
3	175,098	99%	1%	0.60
4	174,819	95%	5%	0.71
5	164,594	94%	6%	0.67
6	161,424	92%	8%	0.71
7	152,338	95%	5%	0.70
8	143,206	93%	7%	0.67
Math				
3	178,085	93%	7%	0.77
4	176,677	94%	6%	0.78
5	166,570	97%	3%	0.78
6	161,968	95%	5%	0.79
7	142,910	97%	3%	0.80
8	107,669	97%	3%	0.82

*Note: Decision consistency was calculated for PBT students only as item parameters were disproportionately based on PBT.

7.3.2. Accuracy

Table 7.23 presents the results of classification accuracy for ELA and Mathematics across all grades. Included in the table are case counts (n-count) and classification accuracy (Accuracy) for all performance levels (All Cuts) and for the Level III cut score. By definition, accuracy associated with the Level III cut is at least as great as that with the entire set of cut scores because there are only two categories for the former, as opposed to the latter, which has four.

For ELA, the estimated accuracy rates indicate that the categorization of a student’s observed performance is in agreement with the location of his or her underlying proficiency from 78% to 80% of the time across all performance levels and 94% to 99% of the time in regard to the Level III cut score. For mathematics, the estimated accuracy rates indicate that the categorization of a student’s observed performance is in agreement with the location of his or her true proficiency from 82% to 86% of the time across all performance levels and 95% to 98% of the time in regard to the Level III cut score.

Table 7.23. Decision Agreement (Accuracy) Estimates*

Grade	N-Count	Accuracy	
		All Cuts	Level III Cut
ELA			
3	175,098	80%	99%
4	174,819	78%	96%
5	164,594	78%	95%
6	161,424	79%	94%
7	152,338	78%	96%
8	143,206	80%	95%
Math			
3	178,085	82%	95%
4	176,677	84%	96%
5	166,570	86%	98%
6	161,968	83%	96%
7	142,910	85%	98%
8	107,669	85%	98%

*Note: Decision agreement was calculated for PBT students only as item parameters were disproportionally based on PBT.

Section 8: Summary of Operational Test Results

This section summarizes the distribution of scale score results on the NYSTP 2017 Grades 3–8 ELA and Mathematics Tests. These include the scale score means, standard deviations, percentile ranks, and performance level distributions for each grade’s population and specific subgroups. Gender, ethnic identification, NRC, ELL, SWD, and SUA variables were used to calculate the results of subgroups required for federal reporting and test equity purposes for both the ELA and mathematics tests. Additionally, the ELL/SUA subgroup is defined as English language learners who use one or more ELL-related accommodations. The SWD/SUA subgroup is defined as examinees with disabilities who use one or more disability-related accommodation(s). For the mathematics analyses, the test translation language is also indicated. (Recall that the ELA tests are not translated, as they are a measure of mastery of the English language.) ELA and mathematics data include examinees with valid scores from all public, religious and independent, and charter schools. Complete scale score frequency distribution tables for ELA and mathematics are located in Appendix Q.

8.1. Scale Score Distribution Summary

Scale score distribution summary tables for ELA and mathematics are presented and discussed. ELA scale score distributions are described first, followed by mathematics. In the following two subsections, ELA and mathematics scale score and subscore statistics are presented for all grades, and across selected subgroups in each grade level. Use caution when interpreting the statistics for subgroups with small number counts that are included in the scale score summaries.

8.1.1. ELA Scale Score and Subscore Distributions

Table 8.1 shows some key statistics characterizing the distribution of ELA scale scores, while Table 8.2 summarizes the ELA subscores derived from the test in each grade. Tables 8.3–8.8 break down the scale scores by selected subgroups. Some general observations from these tables include: Females outperformed Males; Asian and White students outperformed their peers from other reported ethnic groups; students from Low Needs (as identified by NRC) districts outperformed students from other districts (New York City, Big 4 Cities, Urban/Suburban, Rural, Average Needs, and Charter); and ELL students, SWD, SUA, and SWD/SUA tended to underperform the State population (All Students). This pattern of achievement was consistent across all grades.

Table 8.1. ELA Scale Score Distribution Summary

Grade	N-Count	Scale Score		Percentile Ranks				
		Mean	SD	10 th	25 th	50 th	75 th	90 th
3	181,841	307.83	34.78	261	284	312	331	349
4	181,787	306.48	35.23	261	287	308	331	351
5	170,564	301.62	38.88	249	279	304	330	350
6	167,180	299.27	36.13	250	277	303	325	344
7	157,182	307.59	33.82	264	288	310	331	349
8	149,148	306.57	35.45	261	286	309	330	347

Table 8.2. ELA Subscore Summary

Grade	Subscore	N-Count	Subscore		
			Max	Mean	SD
3	Reading	181,841	25	14.80	5.56
	Writing	181,841	22	9.97	4.87
4	Reading	181,787	25	14.28	5.07
	Writing	181,787	22	11.37	4.88
5	Reading	170,564	35	20.48	6.40
	Writing	170,564	22	11.78	4.98
6	Reading	167,180	35	21.30	7.35
	Writing	167,180	22	13.93	5.20
7	Reading	157,182	35	20.93	6.92
	Writing	157,182	22	14.84	5.14
8	Reading	149,148	35	22.57	6.97
	Writing	149,148	22	15.74	4.85

8.1.1.1. ELA Grade 3

Table 8.3 presents the scale score statistics and n-counts of demographic subgroups for Grade 3. The population scale score mean was 307.83 with a standard deviation of 34.78. Female students tended to outperform male students by around 9 scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students from New York City, Average Needs, and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (322.90). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about two-thirds of a standard deviation below the population mean. The students with disabilities (SWD), students tested under accommodations (SUA), and English language learners (ELL) subgroups scored, on average, about one standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed, scoring about 38 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (312): Female (315), Asian (325), Pacific Islander (315), and White (315) students, those attending schools in Low Needs districts (328), and students attending Charter schools (321).

Table 8.3. ELA Grade 3 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 th	25 th	50 th	75 th	90 th
State	All Students	181,841	307.83	34.78	261	284	312	331	349
Gender	Female	89,304	312.73	34.06	266	291	315	338	353
	Male	92,537	303.10	34.82	256	281	306	328	345
Ethnicity	Asian	18,133	322.90	33.63	277	303	325	345	362
	Black	32,831	300.62	35.40	255	277	303	328	345
	Hispanic	51,203	300.03	33.69	256	277	303	325	341

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 th	25 th	50 th	75 th	90 th
Ethnicity	American Indian	1,199	304.03	34.10	256	281	306	328	345
	Multiracial	4,918	310.23	36.22	261	287	312	334	353
	Pacific Islander	507	311.91	33.30	266	291	315	334	353
	White	73,050	312.67	33.33	270	294	315	334	353
NRC	New York	68,543	308.05	35.97	261	284	309	334	353
	Big 4 Cities	7,839	284.41	36.36	237	256	284	312	331
	Urban/Suburban	14,383	295.30	33.10	251	273	297	320	338
	Rural	10,093	297.99	32.61	256	277	300	321	338
	Average Needs	41,413	308.92	31.96	266	288	312	331	345
	Low Needs	18,045	324.20	28.53	287	309	328	341	358
	Charter	11,832	318.41	30.14	277	300	321	341	353
	Religious and Independent	9,592	305.90	36.46	256	284	309	331	349
SWD	All Codes	27,063	278.19	33.24	237	256	277	300	321
SUA	All Codes	12,853	278.14	32.98	237	256	277	300	321
ELL	ELL=Y	19,606	279.47	30.66	237	261	281	300	320
SWD/SUA	SWD & SUA codes	10,663	274.68	32.33	228	251	273	297	315
ELL/SUA	SUA & ELL codes	1,315	269.44	29.29	228	251	270	287	306

8.1.1.2. ELA Grade 4

Table 8.4 contains Grade 4 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 306.48 with a standard deviation of 35.23. Female students tended to outperform male students by around 10 scale score points. Asian, Multiracial, Pacific Islander and White students' scale score means exceeded the state mean scale score, as did those of students from New York City, Average Needs, and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (323.93). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, about one standard deviation below the mean scale score for the population. English language learners tested under accommodations were the lowest performing subgroup analyzed, scoring about 41 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (308): Female (314), Asian (327), Multiracial (311), Pacific Islander (320), and White (314) students, those from Low Needs districts (324), and those enrolled at Charter (320) and Religious and Independent (311) schools.

Table 8.4. ELA Grade 4 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 th	25 th	50 th	75 th	90 th
State	All Students	181,787	306.48	35.23	261	287	308	331	351
Gender	Female	90,245	311.75	34.34	266	292	314	334	351
	Male	91,542	301.28	35.32	256	278	305	324	343
Ethnicity	Asian	18,755	323.93	34.41	278	305	327	346	367
	Black	32,922	298.23	34.75	251	275	299	321	343
	Hispanic	50,851	298.89	33.40	256	278	302	321	338
	American Indian	1,213	300.93	34.60	256	278	302	324	343
	Multiracial	4,443	309.10	35.81	261	287	311	334	351
	Pacific Islander	562	315.48	31.61	274	299	320	338	356
	White	73,041	310.86	34.48	266	292	314	334	351
NRC	New York	70,105	307.62	35.89	261	287	308	331	351
	Big 4 Cities	7,628	281.37	37.43	228	256	282	308	327
	Urban/Suburban	13,420	293.87	33.50	251	270	295	320	334
	Rural	9,605	295.68	33.46	251	274	299	320	334
	Average Needs	39,196	307.05	32.37	266	289	308	327	346
	Low Needs	17,781	322.67	29.54	287	305	324	343	356
	Charter	9,948	315.05	30.32	274	295	320	334	351
	Religious and Independent	14,000	305.75	37.09	256	287	311	331	346
SWD	All Codes	27,879	276.01	33.53	228	256	278	299	320
SUA	All Codes	13,712	275.61	33.25	228	256	278	299	320
ELL	ELL=Y	16,244	272.64	31.17	228	256	274	295	311
SWD/SUA	SWD & SUA codes	10,676	270.85	32.58	228	251	274	292	311
ELL/SUA	SUA & ELL codes	1,240	265.57	29.21	228	251	266	287	302

8.1.1.3. ELA Grade 5

Table 8.5 provides the scale score summary statistics by key demographic subgroups for Grade 5 students. The population scale score mean was 301.62 with a standard deviation of 38.88. Female students tended to outperform male students by around 12 scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students enrolled in New York City, Average Needs, and Low Needs districts and Charter schools. Across all ethnic groups, Asian students earned the highest mean score (319.53). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, one standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest performing subgroup analyzed, scoring about 46 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (304): Female (310), Asian

(323), Pacific Islander (314), Multiracial (307), and White (310) students, those from Average (307) and Low (323) Needs districts and Charter schools (307).

Table 8.5. ELA Grade 5 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 th	25 th	50 th	75 th	90 th
State	All Students	170,564	301.62	38.88	249	279	304	330	350
Gender	Female	83,996	307.85	37.47	257	285	310	334	355
	Male	86,568	295.58	39.27	245	272	298	323	342
Ethnicity	Asian	17,665	319.53	37.40	272	298	323	346	365
	Black	31,811	292.26	38.34	245	268	295	320	338
	Hispanic	47,495	292.63	36.82	245	268	295	317	338
	American Indian	1,166	295.90	39.59	245	268	298	323	346
	Multiracial	3,915	304.55	39.94	253	279	307	330	355
	Pacific Islander	624	309.37	36.70	265	289	314	334	355
	White	67,888	307.51	38.06	257	285	310	334	350
NRC	New York	67,324	302.57	39.51	253	279	304	330	350
	Big 4 Cities	6,982	273.54	40.61	222	245	272	301	327
	Urban/Suburban	12,734	287.28	36.71	240	265	289	314	334
	Rural	9,188	291.18	36.12	245	268	291	317	338
	Average Needs	37,404	303.80	35.93	257	282	307	327	346
	Low Needs	17,372	320.60	32.17	279	301	323	342	360
	Charter	10,120	306.03	35.03	261	285	307	330	350
	Religious and Independent	9,326	297.21	42.72	240	272	301	327	346
SWD	All Codes	27,869	267.45	35.91	222	245	268	291	310
SUA	All Codes	14,483	267.95	36.28	222	245	268	291	314
ELL	ELL=Y	13,920	260.34	32.66	214	240	261	282	301
SWD/SUA	SWD & SUA codes	11,535	262.71	34.88	214	240	265	285	304
ELL/SUA	SUA & ELL codes	1,287	255.38	30.73	214	240	257	275	291

8.1.1.4. ELA Grade 6

Table 8.6 contains Grade 6 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 299.27 with a standard deviation of 36.13. Female students tended to outperform male students by around 12 scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students enrolled in New York City, Average Needs, and Low Needs districts and Charter and Religious and Independent schools. Across ethnic groups, Asian students earned the highest mean score (317.58). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, one standard deviations below the

mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed, scoring about 45 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (303): Female (308), Asian (322), Multiracial (305), Pacific Islander (308), and White (310) students, and those enrolled in Average (305) and Low (320) Needs districts.

Table 8.6. ELA Grade 6 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 th	25 th	50 th	75 th	90 th
State	All Students	167,180	299.27	36.13	250	277	303	325	344
Gender	Female	81,710	305.29	34.46	260	283	308	328	344
	Male	85,470	293.52	36.75	246	269	295	320	338
Ethnicity	Asian	17,434	317.58	34.95	269	298	322	340	356
	Black	32,237	288.58	34.35	243	266	290	313	332
	Hispanic	46,266	290.72	34.09	246	269	293	313	332
	American Indian	1,173	291.79	35.25	246	266	293	316	338
	Multiracial	3,221	303.18	37.72	253	280	305	328	350
	Pacific Islander	495	305.27	34.45	260	283	308	328	350
	White	66,354	305.52	35.35	260	285	310	328	344
NRC	New York	65,146	299.32	36.63	250	275	300	325	344
	Big 4 Cities	6,634	275.40	38.39	230	250	275	303	325
	Urban/Suburban	11,453	284.16	35.10	239	260	285	308	328
	Rural	8,328	293.02	34.87	246	272	295	316	338
	Average Needs	34,828	302.17	34.19	257	283	305	325	344
	Low Needs	16,578	315.93	30.26	277	300	320	338	350
	Charter	10,946	300.89	30.97	260	280	303	322	340
	Religious and Independent	12,997	298.27	37.58	250	280	303	325	340
SWD	All Codes	27,226	266.81	32.18	230	246	268	288	308
SUA	All Codes	14,213	268.65	33.85	225	246	269	293	310
ELL	ELL=Y	12,523	258.31	31.07	217	239	260	280	295
SWD/SUA	SWD & SUA codes	11,001	263.43	32.68	225	243	263	285	305
ELL/SUA	SUA & ELL codes	1,175	253.76	28.26	217	234	255	272	288

8.1.1.5. ELA Grade 7

Table 8.7 presents the Grade 7 scale score statistics and n-counts of demographic subgroups. The population scale score mean was 307.59 with a standard deviation of 33.82. Female students tended to outperform male students by around 11 scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students from New York City, Average and Low Needs districts, and Charter schools. Across ethnic groups, Asian students earned the highest mean score (324.82). Across NRC

categories, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, about one standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed, scoring about 44 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (310): Female (315), Asian (328), Multiracial (315), Pacific Islander (318), and White (315) students as well as those enrolled in Low Needs districts (326), Religious and Independent (311) and Charter (313) schools.

Table 8.7. ELA Grade 7 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 th	25 th	50 th	75 th	90 th
State	All Students	157,182	307.59	33.82	264	288	310	331	349
Gender	Female	76,589	313.50	31.43	274	296	315	334	349
	Male	80,593	301.98	35.04	257	283	305	326	341
Ethnicity	Asian	17,298	324.82	32.08	287	308	328	347	359
	Black	30,419	298.64	31.68	257	280	301	320	334
	Hispanic	42,476	299.65	31.62	257	280	303	320	338
	American Indian	1,163	303.29	31.18	264	288	303	323	341
	Multiracial	2,561	310.74	37.08	263	288	315	338	354
	Pacific Islander	438	315.01	31.66	271	301	318	334	349
	White	62,827	312.45	33.88	270	293	315	334	349
NRC	New York	63,891	309.94	32.95	268	291	310	331	349
	Big 4 Cities	6,265	282.27	36.31	237	261	283	308	326
	Urban/Suburban	10,400	289.86	33.85	245	271	293	313	328
	Rural	8,051	299.08	32.69	257	280	303	320	338
	Average Needs	31,658	307.91	32.94	264	288	310	330	347
	Low Needs	17,177	322.17	28.39	288	305	326	341	354
	Charter	10,329	310.46	27.03	277	296	313	328	341
	Religious and Independent	9,202	304.80	38.34	253	288	311	328	347
SWD	All Codes	25,716	278.63	30.91	240	261	280	300	315
SUA	All Codes	13,724	278.80	32.46	240	259	280	301	318
ELL	ELL=Y	11,460	266.78	31.34	220	249	271	288	303
SWD/SUA	SWD & SUA codes	10,977	274.16	31.25	235	257	277	296	310
ELL/SUA	SUA & ELL codes	1,130	263.08	27.18	228	249	268	280	296

8.1.1.6. ELA Grade 8

Table 8.8 presents the Grade 8 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 306.57 with a standard deviation of 35.45. Female students tended to outperform male students by around 14 scale score points. Asian,

Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students enrolled in New York City, Average and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (324.55). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, one standard deviation below the mean scale score for the population. English language learners tested under accommodations were the lowest performing subgroup analyzed, scoring about 47 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (309), Female (316), Asian (326), Pacific Islander (317), Multiracial (311), and White (316) students, as well as those enrolled in New York City (311) and Low Needs (323) districts, Charter (311), and Religious and Independent (311) schools.

Table 8.8. ELA Grade 8 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 th	25 th	50 th	75 th	90 th
State	All Students	149,148	306.57	35.45	261	286	309	330	347
Gender	Female	72,252	313.81	32.98	273	296	316	334	354
	Male	76,896	299.78	36.33	254	278	304	323	343
Ethnicity	Asian	16,741	324.55	34.02	284	306	326	347	361
	Black	30,368	299.05	33.73	254	281	301	320	338
	Hispanic	40,947	300.13	33.45	257	281	304	323	338
	American Indian	1,180	301.67	34.43	257	284	304	323	343
	Multiracial	2,030	307.53	39.07	254	286	311	334	354
	Pacific Islander	447	311.89	34.71	267	293	317	334	347
	White	57,435	309.93	35.68	264	291	316	334	347
NRC	New York	63,427	309.47	34.36	267	288	311	330	354
	Big 4 Cities	6,117	281.57	39.30	229	257	284	309	330
	Urban/Suburban	9,527	290.26	36.06	243	270	293	316	334
	Rural	7,465	296.95	34.53	254	278	301	320	338
	Average Needs	28,191	305.58	35.65	261	286	309	330	347
	Low Needs	14,518	319.74	30.82	284	304	323	338	354
	Charter	8,456	311.57	26.89	278	296	311	330	343
	Religious and Independent	11,294	306.01	36.93	261	291	311	330	347
SWD	All Codes	23,292	276.33	32.85	234	257	278	298	316
SUA	All Codes	11,739	275.53	34.48	234	254	278	298	317
ELL	ELL=Y	10,406	264.83	31.98	224	247	267	288	301
SWD/SUA	SWD & SUA codes	9,076	270.34	33.36	229	251	273	293	309
ELL/SUA	SUA & ELL codes	732	259.12	30.91	216	243	264	278	296

8.1.2. Mathematics Scale Score Distributions

Table 8.9 shows some key statistics characterizing the distribution of mathematics scale scores, while Table 8.10 summarizes the mathematics subscores derived from the test in each grade. Tables 8.11–8.16 break down the scale scores by selected subgroups. Some general observations from the mathematics data are as follows: Female and Male students performed fairly consistently; Asian students scored considerably higher than other reported ethnic groups; schools belonging to Low Needs districts (as identified by the NRC code) and Charter schools outperformed most other school types (New York City, Big 4 Cities, High Needs Urban/Suburban, and Rural and Average Needs districts). Students taking the Chinese and Korean translations tended to outperform the other translation subgroups (Haitian-Creole, Spanish, and Russian); and ELLs, SWDs, and/or SUAs achieved below the State mean in most percentile ranks. This pattern of achievement was fairly consistent across all grades.

Table 8.9. Mathematics Scale Score Distribution Summary

Grade	N-Count	Scale Score		Percentile Ranks				
		Mean	SD	10 th	25 th	50 th	75 th	90 th
3	183,533	307.41	40.26	254	285	310	335	355
4	183,553	303.90	41.66	250	277	306	333	356
5	171,443	307.08	39.14	256	284	311	334	355
6	167,034	303.01	43.12	245	275	306	333	355
7	155,255	305.82	38.82	246	281	309	334	352
8	116,822	290.24	41.09	229	268	294	318	339

Table 8.10. Mathematics Subscore Summary

Grade	Subscore	N-Count	Subscore		
			Max	Mean	SD
3	Operations and Algebraic Thinking	183,533	25	14.85	6.46
	Number and Operations-Fractions	183,533	11	6.98	2.88
	Measurement and Data	183,533	13	8.16	3.08
4	Operations and Algebraic Thinking	183,553	11	6.94	2.91
	Number and Operations in Base 10	183,553	16	10.54	4.32
	Number and Operations-Fractions	183,553	17	9.43	4.68
5	Number and Operations in Base 10	171,443	17	8.60	4.58
	Number and Operations-Fractions	171,443	23	11.29	6.26
	Measurement and Data	171,443	16	9.72	3.93
6	Ratios and Proportional Relationships	167,034	17	7.93	4.58
	The Number System	167,034	15	8.85	3.42
	Expressions and Equations	167,034	26	14.17	6.79
7	Ratios and Proportional Relationships	155,255	18	7.42	5.08
7	The Number System	155,255	15	7.19	3.92

Grade	Subscore	N-Count	Subscore		
			Max	Mean	SD
	Expressions and Equations	155,255	20	9.53	5.55
8	Expressions and Equations	116,822	29	10.59	6.91
	Functions	116,822	18	8.59	4.58
	Geometry	116,822	16	6.28	3.59

8.1.2.1. Mathematics Grade 3

Table 8.11 presents the Grade 3 scale score statistics and n-counts of demographic subgroups. The population scale score mean was 307.41 with a standard deviation of 40.26. Female and Male students tended to perform similarly. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the state mean scale score, as did those of students from Average and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (330.91). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about two-thirds of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, 0.77 standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 43 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (310): Asian (332), Pacific Islander (315), and White (317) students, as well as those enrolled at Average (314) and Low (326) Needs districts and Charter schools (326). In terms of the 50th-percentile ranks for students using translated forms, they ranged from 266 (Haitian-Creole, n = 122) to 324 (Chinese, n = 705).

Table 8.11. Mathematics Grade 3 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 th	25 th	50 th	75 th	90 th
State	All Students	183,533	307.41	40.26	254	285	310	335	355
Gender	Female	89,809	308.10	39.15	258	285	310	335	355
	Male	93,724	306.74	41.28	250	281	310	335	355
Ethnicity	Asian	18,839	330.91	37.26	285	308	332	355	381
	Black	32,790	296.11	40.73	245	269	297	324	346
	Hispanic	52,159	297.03	38.42	245	272	297	321	346
	American Indian	1,220	302.91	38.29	254	275	304	329	350
	Multiracial	4,874	309.13	41.61	254	285	310	340	362
	Pacific Islander	521	313.38	38.75	258	290	315	335	362
	White	73,130	313.74	38.02	262	292	317	340	362
NRC	New York	70,170	306.50	40.78	254	281	308	335	362
	Big 4 Cities	7,994	281.03	41.69	225	250	281	310	335
	Urban/Suburban	14,524	292.69	38.72	239	266	295	319	342
	Rural	10,030	299.52	38.59	250	275	302	324	346
NRC	Average Needs	40,973	309.80	37.27	262	288	314	335	355

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 th	25 th	50 th	75 th	90 th
	Low Needs	18,126	326.04	34.33	285	306	326	350	370
	Charter	11,820	323.78	37.79	275	299	326	350	370
	Religious and Independent	9,791	301.11	39.46	249	278	304	326	350
SWD	All Codes	27,238	276.23	40.16	225	250	278	304	326
SUA	All Codes	12,450	271.09	40.10	217	245	272	299	321
ELL	ELL=Y	22,278	281.82	37.55	233	258	285	306	329
SWD/SUA	SWD & SUA codes	10,421	267.01	39.50	217	239	266	295	317
ELL/SUA	SUA & ELL codes	1,216	264.18	34.69	217	239	266	288	310
ELL Test Language	Chinese	705	324.14	37.99	275	302	324	350	370
	English	178,571	308.10	40.03	254	285	310	335	355
	Haitian-Creole	122	265.36	37.12	217	239	266	290	317
	Korean	55	316.91	37.81	275	293	308	346	362
	Russian	157	303.01	37.83	245	278	302	324	355
	Spanish	3,923	274.31	35.78	225	250	275	299	317
ELL Test Language	All Translations	4,962	282.55	40.55	233	254	285	308	335

8.1.2.2. Mathematics Grade 4

Table 8.12 presents the Grade 4 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 303.90 with a standard deviation of 41.66. Female and Male students tended to perform similarly. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students enrolled in Average and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (329.52). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, 0.82 standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 47 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (306): Asian (330), Multiracial (310), Pacific Islander (316), and White (314) students, and those enrolled in Average (310) and Low (325) Needs districts and Charter schools (319). In terms of the 50th percentile ranks for students using translated forms, they ranged from: 257 (Haitian-Creole, n = 141) to 323 (Chinese, n = 679).

Table 8.12. Mathematics Grade 4 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 th	25 th	50 th	75 th	90 th
State	All Students	183,553	303.90	41.66	250	277	306	333	356
Gender	Female	90,591	304.00	40.65	254	279	306	330	356
	Male	92,962	303.81	42.62	245	277	306	333	356
Ethnicity	Asian	19,466	329.52	39.20	279	306	330	356	381
	Black	32,924	288.77	40.92	236	264	288	316	341
	Hispanic	51,833	292.61	39.31	241	269	294	319	343
	American Indian	1,229	296.75	40.28	245	272	298	321	347
	Multiracial	4,381	307.24	42.86	254	279	310	336	362
	Pacific Islander	577	313.68	37.94	267	290	316	341	362
	White	73,143	311.74	38.95	264	290	314	336	356
NRC	New York	71,594	302.76	42.70	250	275	302	330	356
	Big 4 Cities	7,786	272.92	43.55	214	241	275	302	328
	Urban/Suburban	13,567	287.36	40.65	236	261	290	316	336
	Rural	9,554	297.50	38.73	245	275	300	323	343
	Average Needs	39,053	308.30	37.67	261	286	310	333	356
	Low Needs	17,976	325.35	33.85	284	306	325	347	370
	Charter	9,924	317.40	39.17	267	290	319	343	370
	Religious and Independent	13,994	298.05	39.62	245	275	300	323	347
SWD	All Codes	27,919	269.95	40.70	214	241	269	296	321
SUA	All Codes	13,923	267.99	41.31	214	241	269	296	321
ELL	ELL=Y	18,970	270.71	38.96	214	245	272	296	319
SWD/SUA	SWD & SUA codes	11,163	262.38	40.64	206	236	264	290	314
ELL/SUA	SUA & ELL codes	1,362	256.58	37.77	206	230	257	284	302
ELL Test Language	Chinese	679	320.65	34.25	275	298	323	343	362
	English	179,065	304.66	41.32	250	279	306	333	356
	Haitian-Creole	141	256.94	42.72	198	230	257	286	319
	Korean	75	318.91	37.87	264	290	323	343	362
	Russian	132	297.83	41.55	241	272	298	325	347
	Spanish	3,461	263.12	38.35	206	236	264	290	312
ELL Test Language	All Translations	4,488	273.58	43.84	214	241	275	304	330

8.1.2.3. Mathematics Grade 5

Table 8.13 presents the Grade 5 demographic subgroup n-counts and scale score statistics. The population scale score mean was 307.08 with a standard deviation of 39.14. Female and male

students tended to perform similarly. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students from Average and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (332.75). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, about 0.84 standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 47 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (311): Asian (337), Multiracial (313), Pacific Islander (320), and White (319) students, as well as those enrolled at Average (316) and Low (331) Needs districts and Charter schools (314). In terms of the 50th percentile ranks for students using translated forms, they ranged from: 259 (Haitian-Creole, n = 116) to 327 (Chinese, n = 686).

Table 8.13. Mathematics Grade 5 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 th	25 th	50 th	75 th	90 th
State	All Students	171,443	307.08	39.14	256	284	311	334	355
Gender	Female	83,994	307.53	37.74	261	284	309	332	352
	Male	87,449	306.65	40.43	250	282	311	334	355
Ethnicity	Asian	18,243	332.75	36.29	287	313	337	355	375
	Black	31,670	292.30	37.81	242	269	296	319	339
	Hispanic	48,166	296.34	37.04	250	273	298	321	341
	American Indian	1,176	301.10	39.46	250	276	304	329	349
	Multiracial	3,691	309.40	40.38	256	284	313	337	359
	Pacific Islander	636	316.15	37.24	269	293	320	341	359
	White	67,861	314.60	36.43	265	296	319	339	355
NRC	New York	68,546	306.46	40.03	256	282	307	334	355
	Big 4 Cities	7,081	278.40	40.60	223	250	279	307	331
	Urban/Suburban	12,746	289.32	38.22	232	265	291	316	334
	Rural	9,019	299.48	36.42	250	279	304	325	341
	Average Needs	36,846	311.82	35.45	265	291	316	334	352
	Low Needs	17,394	328.48	31.75	291	311	331	349	363
	Charter	10,049	313.32	34.61	269	291	314	337	355
	Religious and Independent	9,650	300.27	38.55	250	276	304	327	346
SWD	All Codes	27,573	274.28	37.83	223	250	276	300	321
SUA	All Codes	13,746	273.46	38.76	223	250	276	302	323
ELL	ELL=Y	16,297	274.80	36.54	223	250	276	300	320
SWD/SUA	SWD & SUA codes	11,018	267.39	37.49	215	242	269	294	316
ELL/SUA	SUA & ELL codes	1,226	260.01	34.38	215	232	261	284	304
ELL Test	Chinese	686	323.39	32.08	284	305	327	346	359

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 th	25 th	50 th	75 th	90 th
Language	English	167,266	307.83	38.81	256	284	311	334	355
	Haitian-Creole	116	259.52	35.63	207	237	259	285	309
	Korean	22	329.41	35.11	294	304	326	359	385
	Russian	119	303.13	34.76	256	282	304	327	346
	Spanish	3,234	266.56	34.36	215	242	269	291	309
ELL Test Language	All Translations	4,177	277.07	40.48	223	250	276	304	331

8.1.2.4. Mathematics Grade 6

Table 8.14 presents the Grade 6 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 303.01 with a standard deviation of 43.12. Female students tended to outperform male students by around 4 scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students enrolled in Average and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (331.25). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, 0.84 standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 49 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (306): Female (308), Asian (335), Multiracial (312), Pacific Islander (312), and White (318) students, as well as those enrolled in Average (314) and Low (331) Needs districts and Charter schools (308). In terms of the 50th percentile ranks for students using translated forms, they ranged from: 268 (Spanish, n = 4,210) to 321 (Chinese, n = 836; and Korean, n = 35).

Table 8.14. Mathematics Grade 6 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 th	25 th	50 th	75 th	90 th
State	All Students	167,034	303.01	43.12	245	275	306	333	355
Gender	Female	81,345	305.13	41.70	251	278	308	333	355
	Male	85,689	300.99	44.34	238	272	304	331	355
Ethnicity	Asian	18,038	331.25	41.71	278	306	335	358	381
	Black	31,860	284.66	41.06	230	256	286	314	335
	Hispanic	46,735	290.06	39.16	238	265	291	318	340
	American Indian	1,172	293.39	42.74	238	265	295	321	349
	Multiracial	3,116	308.14	44.34	251	281	312	340	361
	Pacific Islander	508	307.70	44.02	251	284	312	337	358
	White	65,605	313.26	39.92	261	291	318	340	362
NRC	New York	66,463	300.28	44.96	245	268	300	331	358

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 th	25 th	50 th	75 th	90 th
	Big 4 Cities	6,649	274.42	43.88	213	245	275	306	331
	Urban/Suburban	11,262	282.25	39.57	230	256	284	310	331
	Rural	8,120	297.08	38.52	245	275	300	323	344
	Average Needs	34,087	309.92	38.86	261	289	314	335	355
	Low Needs	16,406	328.91	35.43	286	308	331	352	370
	Charter	10,836	306.40	39.18	256	281	308	333	355
	Religious and Independent	13,104	299.92	39.84	245	275	302	327	347
SWD	All Codes	26,644	264.99	38.93	213	238	265	291	314
SUA	All Codes	13,180	267.05	40.62	213	238	268	295	318
ELL	ELL=Y	14,865	268.05	38.05	221	245	268	291	318
SWD/SUA	SWD & SUA codes	10,370	260.94	39.21	213	230	261	289	310
ELL/SUA	SUA & ELL codes	1,065	253.97	32.15	213	230	256	275	293
ELL Test Language	Chinese	839	319.67	39.46	268	298	321	347	370
	English	161,687	303.84	43.06	245	278	306	333	355
	Haitian-Creole	129	271.42	29.17	238	256	272	289	308
	Korean	35	311.94	45.88	251	281	321	340	366
	Russian	134	307.11	36.67	265	284	308	329	358
	Spanish	4,210	268.39	29.28	230	251	268	286	308
ELL Test Language	All Translations	5,347	277.77	36.90	238	251	272	300	329

8.1.2.5. Mathematics Grade 7

Table 8.15 presents the Grade 7 n-counts and scale score statistics for key demographic subgroups. The population scale score mean was 305.82 with a standard deviation of 38.82. Female students tended to outperform male students by around 4 scale score points. Asian, Multiracial, Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students from Average and Low Needs districts and Charter schools. Across ethnic groups, Asian students earned the highest mean score (332.68). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score—by about three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, 0.85 standard deviations below the mean scale score for the population. English language learners tested under accommodations were the lowest-performing subgroup analyzed for English forms, scoring about 48 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (309): Female (311), Asian (338), Multiracial (315), Pacific Islander (318), and White (319) students, those enrolled in Average (315) and Low (331) Needs districts, and Charter schools (312). In terms of the 50th percentile ranks for students using translated forms, they ranged from: 277 (Haitian-Creole, n = 104) to 333 (Korean, n = 865).

Table 8.15. Mathematics Grade 7 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 th	25 th	50 th	75 th	90 th
State	All Students	155,255	305.82	38.82	246	281	309	334	352
Gender	Female	75,358	307.65	37.27	259	285	311	334	352
	Male	79,897	304.09	40.15	246	277	308	334	352
Ethnicity	Asian	17,576	332.68	35.66	288	312	338	357	377
	Black	29,690	289.87	36.97	238	267	293	315	337
	Hispanic	42,595	293.33	36.51	238	273	295	319	338
	American Indian	1,149	297.38	36.53	246	277	299	322	340
	Multiracial	2,429	311.23	39.45	259	285	315	341	360
	Pacific Islander	451	311.60	38.53	259	293	318	337	357
	White	61,365	314.41	35.35	267	295	319	338	354
NRC	New York	64,725	305.10	40.70	246	281	306	334	357
	Big 4 Cities	6,221	275.70	37.36	230	246	277	303	325
	Urban/Suburban	9,993	284.69	36.03	230	259	288	309	330
	Rural	7,699	298.44	34.98	246	277	301	323	340
	Average Needs	30,298	310.14	34.55	267	291	315	334	350
	Low Needs	16,467	327.15	30.73	291	312	331	346	360
	Charter	10,180	309.67	34.85	267	288	312	334	352
	Religious and Independent	9,545	304.13	36.24	246	285	308	330	346
SWD	All Codes	24,834	271.85	34.78	223	246	273	295	317
SUA	All Codes	11,803	273.57	36.31	223	246	277	299	321
ELL	ELL=Y	13,913	272.56	35.84	223	246	273	295	318
SWD/SUA	SWD & SUA codes	9,455	268.00	34.66	223	238	273	291	312
ELL/SUA	SUA & ELL codes	888	258.18	29.56	223	234	259	281	297
ELL Test Language	Chinese	865	328.01	32.87	285	311	333	350	367
	English	150,075	306.74	38.45	246	285	309	334	352
	Haitian-Creole	140	273.39	34.17	230	246	277	298	316
	Korean	39	321.15	30.97	281	297	326	345	360
	Russian	147	299.74	35.44	259	281	297	323	341
	Spanish	3,989	267.50	32.46	223	246	273	291	309
ELL Test Language	All Translations	5,180	279.08	39.94	230	246	277	305	334

8.1.2.6. Mathematics Grade 8

Table 8.16 presents the Grade 8 scale score statistics and n-counts for key demographic subgroups. The population scale score mean was 290.24 with a standard deviation of 41.09. Female students tended to outperform male students by around 8 scale score points. Asian,

Pacific Islander, and White students' scale score means exceeded the State mean scale score, as did those of students enrolled in New York City, Average and Low Needs districts and Charter and Religious and Independent schools. Across ethnic groups, Asian students earned the highest mean score (319.97). Across NRC categories, students from Big 4 Cities districts earned the lowest mean score—by three-quarters of a standard deviation below the population mean. The SWD, SUA, and ELL subgroups scored, on average, about three-quarters of a standard deviation below the mean scale score for the population. English language learners tested under accommodations were the lowest performing subgroup analyzed for English forms, scoring about 42 scale score points below the State mean. At the 50th percentile, the following groups exceeded that of the population (294): Female (298), Asian (324), Pacific Islander (300), and White (302) students, as well as those enrolled in Average (296) and Low (314) Needs districts and Charter (307) and Religious and Independent (300) schools. In terms of the 50th percentile ranks for students using translated forms, they ranged from: 268 (Spanish, n = 3,812; and Haitian-Creole, n = 129) to 333 (Korean, n = 20).

Table 8.16. Mathematics Grade 8 Scale Score Distribution by Subgroup

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 th	25 th	50 th	75 th	90 th
State	All Students	116,822	290.24	41.09	229	268	294	318	339
Gender	Female	55,480	294.65	39.28	244	273	298	321	341
	Male	61,342	286.26	42.28	221	262	292	317	336
Ethnicity	Asian	11,441	319.97	40.28	268	296	324	346	370
	Black	25,291	277.77	40.00	221	254	280	306	327
	Hispanic	35,720	283.47	38.92	229	262	287	311	330
	American Indian	948	282.82	38.68	229	262	284	309	333
	Multiracial	1,497	289.93	43.08	229	262	294	319	341
	Pacific Islander	354	295.88	41.36	229	268	300	324	344
	White	41,571	295.61	38.75	244	277	302	321	339
NRC	New York	53,231	292.21	42.11	229	268	294	321	344
	Big 4 Cities	5,546	262.95	40.78	213	229	262	292	315
	Urban/Suburban	7,595	268.37	37.20	213	244	273	296	312
	Rural	5,964	281.40	37.85	221	262	287	307	324
	Average Needs	18,802	289.41	36.44	229	273	296	315	330
	Low Needs	7,875	308.62	34.29	268	294	314	331	346
	Charter	6,573	303.91	37.54	254	280	307	330	349
	Religious and Independent	11,163	294.76	40.54	229	273	300	322	341
SWD	All Codes	21,096	260.64	37.59	213	229	262	287	307
SUA	All Codes	9,780	260.46	39.29	205	229	262	289	311
ELL	ELL=Y	12,327	266.63	38.80	213	244	268	292	315
SWD/SUA	SWD & SUA codes	7,776	255.63	38.17	205	229	262	284	304

Demographic Category		N-Count	Scale Score		Percentile Ranks				
			Mean	SD	10 th	25 th	50 th	75 th	90 th
ELL/SUA	SUA & ELL codes	641	248.29	33.09	205	221	254	273	289
ELL Test Language	Chinese	755	322.14	38.87	273	298	327	349	365
	English	111,967	290.96	40.96	229	268	296	319	339
	Haitian-Creole	129	265.40	35.57	213	244	268	292	306
	Korean	20	328.35	35.76	267	323	333	348	367
	Russian	139	294.12	37.73	244	273	296	319	341
	Spanish	3,812	263.54	33.36	221	244	268	287	304
ELL Test Language	All Translations	4,855	273.85	40.74	221	244	273	298	328

8.2. Performance Level Distribution Summary

Students are classified as NYS Level I, NYS Level II, NYS Level III, or NYS Level IV. The cut scores were established in 2013 during the standard-setting. Table 6.13 and Table 6.14 show the ELA and Mathematics cut scores, respectively, used for classification of students into the four performance-level categories in 2017. It is inappropriate to compare scale scores across grades as they neither measure the same content, nor are they on the same scale. During the standard-setting process, while cut scores were set separately for different grades within a subject, additional care was taken to vertically articulate performance levels; see Section 8 and Appendix P in the 2013 technical report (NYSED, 2014) for details. While vertical articulation helps to build consistent meaning to the performance levels, the very nature of grade-specific content, differing performance expectations, and panel-set cut scores result in cut score differences across grades.

8.2.1. ELA Test Performance Level Distributions

Table 8.17 shows the performance level distribution for all examinees from public, charter, and religious and independent schools with valid ELA scores. Performance level data for selected subgroups of students were also examined. In general, these distributions reflect the same achievement trends in the scale score summary discussion. Across Tables 8.18 through 8.23, more Female students were classified in Level III and above categories than were Male students. Similarly, more Asian and White students were classified in Level III and above categories than were their peers from other reported ethnic groups. Consistent with the pattern shown in scale score distribution across the subgroups, students from Low and Average Needs districts outperformed students from High Needs districts (New York City, Big 4 Cities, Urban/Suburban, and Rural). The Level III and above rates for students in the ELL, SWD, and SUA subgroups were low, compared to the total population of examinees.

Table 8.17. ELA Test Performance Level Distributions

Grade	N-Count	Performance Levels				
		Level I	Level II	Level III	Level IV	Level III & IV
3	181,841	27.93	29.16	35.87	7.03	42.90
4	181,787	24.14	34.59	25.43	15.84	41.27

Grade	N-Count	Performance Levels				
		Level I	Level II	Level III	Level IV	Level III & IV
5	170,564	32.95	31.81	22.33	12.92	35.25
6	167,180	28.91	38.77	16.31	16.00	32.31
7	157,182	21.72	36.35	29.15	12.78	41.93
8	149,148	21.10	33.32	30.61	14.97	45.58

8.2.1.1. ELA Grade 3

Table 8.18 presents the ELA Grade 3 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 43% of students achieved Level III and Level IV. About 48% of Female students were at Level III or above, as compared to 37% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (60%) students and students from Low Needs districts (65%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 20–34% of students in those same performance categories. Only about 12% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (43%), Female (48%), Asian (61%), Multiracial (46%), Pacific Islander (49%), White (49%) students, and those enrolled in Average (43%) and Low (65%) Needs districts and Charter (56%) schools.

Table 8.18. ELA Grade 3 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	181,841	27.93	29.16	35.87	7.03	42.90
Gender	Female	89,304	23.46	28.12	39.24	9.18	48.43
	Male	92,537	32.25	30.17	32.61	4.96	37.57
Ethnicity	Asian	18,133	15.04	24.04	45.30	15.61	60.92
	Black	32,831	35.89	29.55	29.74	4.82	34.56
	Hispanic	51,203	35.65	31.43	29.02	3.89	32.92
	American Indian	1,199	33.19	28.86	32.28	5.67	37.95
	Multiracial	4,918	26.41	27.67	36.70	9.21	45.91
	Pacific Islander	507	22.29	28.80	40.43	8.48	48.92
	White	73,050	22.20	28.78	41.06	7.96	49.02
NRC	New York	68,543	28.85	28.64	33.92	8.59	42.51
	Big 4 Cities	7,839	54.41	25.60	17.87	2.12	19.99
	Urban/Suburban	14,383	41.19	31.89	24.34	2.57	26.91
	Rural	10,093	37.73	32.53	26.98	2.76	29.74
	Average Needs	41,413	25.00	31.81	37.70	5.49	43.18
	Low Needs	18,045	10.70	24.10	53.74	11.46	65.20
	Charter	11,832	15.98	27.89	46.94	9.19	56.13

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
NCR	Religious and Independent	9,592	29.37	27.98	35.96	6.69	42.65
SWD	All Codes	27,063	62.38	25.13	11.68	0.81	12.49
SUA	All Codes	12,853	62.23	25.48	11.58	0.71	12.29
ELL	ELL=Y	19,606	60.88	28.38	10.34	0.39	10.74
SWD/ SUA	SWD & SUA codes	10,663	66.66	23.60	9.27	0.47	9.73
ELL/SUA	SUA & ELL	1,315	75.36	18.94	5.63	0.08	5.70

8.2.1.2. ELA Grade 4

Table 8.19 presents the ELA Grade 4 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 41% of students achieved Level III and Level IV. About 47% of Female students were at Level III or above, as compared to 35% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (62%) students and students from Low Needs districts (63%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 18–32% of students in those same performance categories. Only about 9% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (41%): Female (47%), Asian (62%), Multiracial (44%), Pacific Islander (51%), and White (47%) students as well as those enrolled in New York City (42%) and Low (63%) Needs districts, and Charter schools (52%).

Table 8.19. ELA Grade 4 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	181,787	24.14	34.59	25.43	15.84	41.27
Gender	Female	90,245	19.51	33.34	27.53	19.63	47.16
	Male	91,542	28.70	35.83	23.37	12.09	35.46
Ethnicity	Asian	18,755	11.76	25.83	30.14	32.27	62.41
	Black	32,922	31.99	36.16	21.80	10.05	31.85
	Hispanic	50,851	30.35	38.31	21.83	9.50	31.34
	American Indian	1,213	29.27	36.36	23.25	11.13	34.38
	Multiracial	4,443	23.21	32.68	25.03	19.09	44.11
	Pacific Islander	562	14.77	33.63	30.60	21.00	51.60
	White	73,041	19.50	33.64	28.39	18.47	46.86
NRC	New York	70,105	23.97	34.10	24.37	17.56	41.93
	Big 4 Cities	7,628	52.58	29.76	12.00	5.66	17.66
	Urban/Suburban	13,420	36.54	37.04	19.32	7.10	26.42
	Rural	9,605	33.78	39.10	19.28	7.83	27.11

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
NRC	Average Needs	39,196	21.73	37.81	26.56	13.90	40.47
	Low Needs	17,781	8.91	28.57	35.93	26.59	62.52
	Charter	9,948	14.92	32.93	32.35	19.80	52.15
	Religious and Independent	14,000	23.53	34.19	26.76	15.52	42.29
SWD	All Codes	27,879	58.21	30.71	8.78	2.30	11.08
SUA	All Codes	13,712	57.84	31.99	8.31	1.86	10.17
ELL	ELL=Y	16,244	61.51	32.06	5.75	0.68	6.43
SWD/ SUA	SWD & SUA codes	10,676	63.89	28.88	6.07	1.16	7.23
ELL/SUA	SUA & ELL codes	1,240	72.18	25.24	2.42	0.16	2.58

8.2.1.3. ELA Grade 5

Table 8.20 presents the ELA Grade 5 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 35% of students achieved Level III and Level IV. About 41% of Female students were at Level III or above, as compared to 29% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (55%) students and students from Low Needs districts (56%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 14–25% of students in those same performance categories. Only about 6% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (35%): Female (41%), Asian (55%), Multiracial (39%), Pacific Islander (43%), and White (42%) students, as well as those enrolled in New York City (36%), Average (36%), and Low (56%) Needs districts and Charter schools (38%).

Table 8.20. ELA Grade 5 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	170,564	32.95	31.81	22.33	12.92	35.25
Gender	Female	83,996	27.17	31.72	24.70	16.41	41.11
	Male	86,568	38.55	31.88	20.02	9.54	29.57
Ethnicity	Asian	17,665	17.11	27.66	29.49	25.74	55.23
	Black	31,811	42.68	31.69	17.62	8.01	25.62
	Hispanic	47,495	41.69	33.43	17.69	7.18	24.88
	American Indian	1,166	37.82	32.25	19.73	10.21	29.93
	Multiracial	3,915	30.68	30.50	22.78	16.04	38.83
	Pacific Islander	624	24.52	32.37	28.04	15.06	43.11
	White	67,888	26.51	31.86	25.87	15.76	41.63
NRC	New York	67,324	33.06	30.90	21.66	14.38	36.04

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
NRC	Big 4 Cities	6,982	62.37	23.52	10.00	4.11	14.11
	Urban/Suburban	12,734	47.86	31.69	15.16	5.29	20.45
	Rural	9,188	43.49	33.37	16.97	6.17	23.14
	Average Needs	37,404	29.38	34.77	23.94	11.91	35.85
	Low Needs	17,372	13.78	29.96	33.40	22.86	56.26
	Charter	10,120	27.35	34.76	24.79	13.09	37.89
	Religious and Independent	9,326	35.16	31.55	21.80	11.49	33.29
SWD	All Codes	27,869	70.44	22.39	5.71	1.45	7.17
SUA	All Codes	14,483	69.67	22.98	5.98	1.37	7.35
ELL	ELL=Y	13,920	78.93	18.72	2.10	0.24	2.35
SWD/ SUA	SWD & SUA codes	11,535	75.46	20.01	4.02	0.51	4.53
ELL/SUA	SUA & ELL codes	1,287	85.78	13.05	1.01	0.16	1.17

8.2.1.4. ELA Grade 6

Table 8.21 presents the ELA Grade 6 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 32% of students achieved Level III and Level IV. About 38% of Female students were at Level III or above, as compared to 27% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (55%) students and students from Low Needs districts (51%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 15–21% of students in those same performance categories. Only about 5% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (32%): Female (38%), Asian (55%), Multiracial (38%), Pacific Islander (39%), and White (39%) students, as well as those from Average (34%) and Low (51%) Needs districts.

Table 8.21. ELA Grade 6 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	167,180	28.91	38.77	16.31	16.00	32.31
Gender	Female	81,710	23.05	39.01	18.06	19.88	37.94
	Male	85,470	34.52	38.55	14.65	12.29	26.94
Ethnicity	Asian	17,434	14.47	30.24	22.19	33.10	55.29
	Black	32,237	39.93	39.82	12.09	8.16	20.25
	Hispanic	46,266	36.88	41.21	12.83	9.07	21.91
	American Indian	1,173	39.13	36.40	13.64	10.83	24.47

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
Ethnicity	Multiracial	3,221	25.95	36.36	16.95	20.74	37.69
	Pacific Islander	495	24.44	36.57	19.60	19.39	38.99
	White	66,354	21.80	38.98	19.24	19.98	39.22
NRC	New York	65,146	30.39	37.34	15.12	17.15	32.27
	Big 4 Cities	6,634	55.49	29.86	8.83	5.82	14.65
	Urban/Suburban	11,453	44.64	38.32	10.19	6.85	17.03
	Rural	8,328	34.79	40.55	14.41	10.25	24.66
	Average Needs	34,828	24.83	41.05	17.62	16.50	34.12
	Low Needs	16,578	11.52	37.36	24.34	26.78	51.12
	Charter	10,946	25.14	43.66	17.91	13.29	31.20
	Religious and Independent	12,997	26.54	41.37	17.62	14.46	32.08
SWD	All Codes	27,226	66.88	28.03	3.67	1.42	5.09
SUA	All Codes	14,213	63.57	29.90	4.64	1.89	6.54
ELL	ELL=Y	12,523	76.95	21.12	1.52	0.41	1.92
SWD/ SUA	SWD & SUA codes	11,001	70.18	25.90	2.95	0.98	3.93
ELL/SUA	SUA & ELL codes	1,175	84.94	14.30	0.68	0.09	0.77

8.2.1.5. ELA Grade 7

Table 8.22 presents the ELA Grade 7 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 42% of students achieved Level III and Level IV. About 48% of Female students were at Level III or above, as compared to 36% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (65%) students and students from Low Needs (62%) districts. The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 17–31% of students in those same performance categories. Only about 8% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (42%): Female (48%), Asian (65%), Multiracial (48%), Pacific Islander (54%), and White (49%) students, as well as those enrolled in New York City (43%), Average (43%), and Low (62%) Needs districts, Religious and Independent (43%) and Charter (43%) schools.

Table 8.22. ELA Grade 7 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	157,182	21.72	36.35	29.15	12.78	41.93
Gender	Female	76,589	16.12	35.51	32.46	15.91	48.37
	Male	80,593	27.04	37.14	26.01	9.80	35.81

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
Ethnicity	Asian	17,298	9.92	24.66	37.23	28.19	65.42
	Black	30,419	28.76	42.06	23.09	6.09	29.18
	Hispanic	42,476	27.72	41.76	24.26	6.27	30.53
	American Indian	1,163	23.04	43.94	24.08	8.94	33.02
	Multiracial	2,561	21.94	29.87	29.95	18.24	48.18
	Pacific Islander	438	15.75	30.14	38.13	15.98	54.11
	White	62,827	17.51	33.32	33.17	16.00	49.17
NRC	New York	63,891	19.74	37.02	28.45	14.80	43.25
	Big 4 Cities	6,265	50.57	32.21	13.79	3.43	17.22
	Urban/Suburban	10,400	40.21	38.88	16.91	4.00	20.91
	Rural	8,051	29.69	39.83	23.59	6.89	30.48
	Average Needs	31,658	21.09	36.38	30.56	11.97	42.52
	Low Needs	17,177	8.75	29.50	40.62	21.13	61.75
	Charter	10,329	15.45	41.31	33.52	9.72	43.24
	Religious and Independent	9,202	21.08	35.86	32.15	10.91	43.06
SWD	All Codes	25,716	54.92	35.90	8.05	1.12	9.17
SUA	All Codes	13,724	54.61	34.70	9.27	1.42	10.69
ELL	ELL=Y	11,460	68.76	28.50	2.51	0.23	2.74
SWD/ SUA	SWD & SUA codes	10,977	60.73	32.57	6.09	0.62	6.70
ELL/SUA	SUA & ELL codes	1,130	79.65	19.47	0.88		0.88

8.2.1.6. ELA Grade 8

Table 8.23 presents the ELA Grade 8 performance level distributions and n-counts of demographic subgroups. Statewide, a combined 46% of students achieved Level III and Level IV. About 53% of Female students were at Level III or above, as compared to 38% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (68%) students and students from Low Needs (64%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 21–37% of students in those same performance categories. Only about 9% of the SWD, SUA, and ELL subgroups on average earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (46%): Female (54%), Asian (68%), Multiracial (49%), Pacific Islander (53%), and White (51%) students, as well as those attending New York City (47%) and Low (63%) Needs districts, those enrolled in Charter (49%), and Religious and Independent (47%) schools.

Table 8.23. ELA Grade 8 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	149,148	21.10	33.32	30.61	14.97	45.58
Gender	Female	72,252	14.64	31.83	34.03	19.50	53.53
	Male	76,896	27.17	34.71	27.39	10.72	38.11
Ethnicity	Asian	16,741	9.89	22.33	35.47	32.32	67.79
	Black	30,368	26.50	38.47	26.28	8.74	35.02
	Hispanic	40,947	25.43	37.74	27.98	8.86	36.83
	American Indian	1,180	24.32	39.24	24.75	11.69	36.44
	Multiracial	2,030	22.61	28.62	30.30	18.47	48.77
	Pacific Islander	447	16.78	30.43	32.89	19.91	52.80
	White	57,435	18.35	30.70	33.46	17.48	50.94
NRC	New York	63,427	18.86	33.69	30.46	16.99	47.45
	Big 4 Cities	6,117	47.82	30.65	16.12	5.41	21.53
	Urban/Suburban	9,527	37.06	36.03	20.71	6.19	26.90
	Rural	7,465	29.35	37.43	25.33	7.89	33.22
	Average Needs	28,191	21.89	32.99	30.83	14.29	45.12
	Low Needs	14,518	9.81	26.58	40.02	23.59	63.61
	Charter	8,456	12.44	38.11	37.10	12.35	49.44
	Religious and Independent	11,294	19.18	33.55	33.67	13.60	47.27
SWD	All Codes	23,292	53.80	35.18	9.65	1.38	11.03
SUA	All Codes	11,739	55.06	32.82	10.32	1.80	12.11
ELL	ELL=Y	10,406	68.49	27.63	3.59	0.29	3.88
SWD/ SUA	SWD & SUA codes	9,076	61.84	30.35	6.96	0.84	7.80
ELL/SUA	SUA & ELL codes	732	78.28	19.95	1.78		1.78

8.2.2. Mathematics Test Performance Level Distributions

Table 8.24 shows the performance level distributions for all examinees from public, charter, and religious and independent schools with valid scores, and presents mathematics performance level data for total populations of students in Grades 3–8. Performance level data for selected subgroups of students were also examined. In general, these summaries reflect the same achievement trends as in the scale score summary discussion. Across Table 8.25 through Table 8.30, Male and Female students performed similarly across grades. More White, Pacific Islander, and Asian students were classified in Level III and above, as compared to their peers from other ethnic subgroups. Students from Low and Average Needs districts and Charter schools outperformed students from High Needs districts (New York City, Big 4 Cities, High Needs Urban/Suburban, and High Needs Rural), and Religious and Independent schools. The subgroups that used the Korean or Chinese translations outperformed other test translation subgroups. The

Level III and above rates for SWD and SUA subgroups were low, compared to the total population of examinees. The n-counts for the Haitian-Creole, Korean, and Russian translation subgroups were very low, and the results might have been heavily influenced by very high and/or very low achieving individual students.

Table 8.24. Mathematics Test Performance Level Distributions

Grade	N-Count	Performance Levels				
		Level I	Level II	Level III	Level IV	Level III & IV
3	183,533	24.95	27.04	24.98	23.03	48.01
4	183,553	26.98	30.62	21.92	20.49	42.40
5	171,443	32.49	24.91	27.08	15.52	42.60
6	167,034	29.96	30.68	18.78	20.58	39.36
7	155,255	32.98	29.62	24.01	13.40	37.40
8	116,822	40.82	37.24	15.70	6.24	21.94

8.2.2.1. Mathematics Grade 3

Table 8.25 presents the Mathematics Grade 3 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 48% of students achieved Level III and Level IV. About 48% of both Female and Male students were at Level III or above. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (72%) students and students from Low Needs (69%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 24–36% of students in those same performance categories. Only about 19% of the SWD, SUA, and ELL subgroups, on average, earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (48%): Asian (72%), Multiracial (50%), Pacific Islander (54%), and White (56%) students, as well as those enrolled at Average (51%) and Low (69%) Needs districts and Charter schools (64%). For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 11% (Haitian-Creole) to 66% (Chinese).

Table 8.25. Mathematics Grade 3 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	183,533	24.95	27.04	24.98	23.03	48.01
Gender	Female	89,809	24.12	27.92	25.12	22.85	47.97
	Male	93,724	25.75	26.20	24.85	23.20	48.05
Ethnicity	Asian	18,839	9.42	18.29	27.24	45.05	72.29
	Black	32,790	35.28	29.17	20.07	15.47	35.54
	Hispanic	52,159	33.02	31.17	21.53	14.28	35.81
	American Indian	1,220	28.03	29.26	24.43	18.28	42.70
	Multiracial	4,874	24.62	25.63	24.17	25.58	49.75
	Pacific Islander	521	18.43	27.45	29.56	24.57	54.13

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
	White	73,130	18.58	25.45	29.09	26.88	55.97
NRC	New York	70,170	26.36	27.74	23.18	22.72	45.90
	Big 4 Cities	7,994	50.38	25.92	14.75	8.96	23.71
	Urban/Suburban	14,524	37.62	30.36	20.13	11.89	32.02
	Rural	10,030	30.44	29.53	24.42	15.61	40.03
	Average Needs	40,973	21.05	27.82	28.53	22.60	51.13
	Low Needs	18,126	9.36	21.26	31.65	37.73	69.38
	Charter	11,820	13.27	22.10	26.40	38.23	64.64
	Religious and Independent	9,791	29.00	28.97	25.04	17.00	42.04
SWD	All Codes	27,238	55.33	25.66	12.50	6.51	19.01
SUA	All Codes	12,450	59.90	24.18	10.89	5.03	15.92
ELL	ELL=Y	22,278	49.11	29.94	14.28	6.67	20.95
SWD/ SUA	SWD & SUA codes	10,421	64.09	22.54	9.49	3.88	13.37
ELL/SUA	SUA & ELL codes	1,216	69.24	21.88	6.99	1.89	8.88
ELL Test Language	Chinese	705	12.34	22.13	28.51	37.02	65.53
	English	178,571	24.26	27.04	25.30	23.41	48.71
	Haitian-Creole	122	68.03	21.31	8.20	2.46	10.66
	Korean	55	12.73	38.18	16.36	32.73	49.09
	Russian	157	27.39	32.48	21.02	19.11	40.13
	Spanish	3,923	57.56	27.91	10.73	3.80	14.53
	All Translations	183,533	24.95	27.04	24.98	23.03	48.01

8.2.2.2. Mathematics Grade 4

Table 8.26 presents the Mathematics Grade 4 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 42% of students achieved Level III and Level IV. About 42% of both Female and Male students were at Level III or above. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (68%) students and students from Low Needs (67%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 18–29% of students in those same performance categories. Only about 13% of the SWD, SUA, and ELL subgroups, on average, earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (42%): Male (43%), Asian (68%), Multiracial (47%), Pacific Islander (52%), and White (51%) students, as well as students enrolled in Average (47%) and Low (67%) Needs and Charter schools (55%). For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 9% (Spanish) to 63% (Korean).

Table 8.26. Mathematics Grade 4 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	183,553	26.98	30.62	21.92	20.49	42.40
Gender	Female	90,591	26.62	31.60	21.99	19.79	41.77
	Male	92,962	27.33	29.66	21.85	21.17	43.01
Ethnicity	Asian	19,466	10.18	21.36	25.35	43.11	68.46
	Black	32,924	41.29	31.56	15.83	11.31	27.14
	Hispanic	51,833	35.96	34.59	17.77	11.67	29.44
	American Indian	1,229	33.28	33.60	18.47	14.65	33.12
	Multiracial	4,381	25.36	27.92	23.12	23.60	46.72
	Pacific Islander	577	17.50	30.16	26.52	25.82	52.34
	White	73,143	18.72	29.95	26.62	24.71	51.33
NRC	New York	71,594	29.69	30.33	19.34	20.64	39.98
	Big 4 Cities	7,786	56.24	25.80	11.44	6.51	17.96
	Urban/Suburban	13,567	41.21	31.84	17.09	9.85	26.95
	Rural	9,554	29.65	34.62	22.33	13.40	35.72
	Average Needs	39,053	20.74	32.28	26.03	20.94	46.97
	Low Needs	17,976	8.55	24.94	30.81	35.70	66.51
	Charter	9,924	17.39	27.73	24.13	30.74	54.88
	Religious and Independent	13,994	29.08	35.54	20.87	14.51	35.38
SWD	All Codes	27,919	60.10	25.95	9.36	4.59	13.95
SUA	All Codes	13,923	60.49	26.05	9.55	3.91	13.46
ELL	ELL=Y	18,970	58.39	28.85	9.14	3.63	12.77
SWD/ SUA	SWD & SUA codes	11,163	66.29	23.23	7.70	2.78	10.48
ELL/SUA	SUA & ELL codes	1,362	72.47	21.95	4.41	1.17	5.58
ELL Test Language	Chinese	679	12.37	26.80	31.08	29.75	60.82
	English	179,065	26.25	30.75	22.17	20.83	43.00
	Haitian-Creole	141	69.50	18.44	9.93	2.13	12.06
	Korean	75	20.00	17.33	33.33	29.33	62.67
	Russian	132	33.33	29.55	22.73	14.39	37.12
	Spanish	3,461	65.82	25.17	7.34	1.68	9.01
	All Translations	183,553	26.98	30.62	21.92	20.49	42.40

8.2.2.3. Mathematics Grade 5

Table 8.27 presents the Mathematics Grade 5 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 43% of students achieved Level III and Level

IV. About 43% of both Female and Male students were at Level III or above. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (71%) students and students from Low Needs districts (68%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 18–30% of students in those same performance categories. Only about 12% of the SWD, SUA, and ELL subgroups, on average, earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (43%): Asian (71%), Multiracial (46%), Pacific Islander (53%), and White (52%) students, as well as those enrolled in Average (48%) and Low (68%) Needs districts and Charter schools (47%). For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 5% (Haitian-Creole) to 64% (Chinese).

Table 8.27. Mathematics Grade 5 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	171,443	32.49	24.91	27.08	15.52	42.60
Gender	Female	83,994	31.77	26.29	27.03	14.92	41.94
	Male	87,449	33.17	23.59	27.14	16.10	43.23
Ethnicity	Asian	18,243	12.69	16.67	32.05	38.59	70.64
	Black	31,670	47.41	26.82	18.84	6.93	25.77
	Hispanic	48,166	43.07	27.17	21.80	7.96	29.76
	American Indian	1,176	38.69	25.26	23.64	12.41	36.05
	Multiracial	3,691	31.54	22.92	27.15	18.40	45.54
	Pacific Islander	636	25.00	22.01	32.55	20.44	52.99
	White	67,861	23.34	24.77	33.35	18.54	51.89
NRC	New York	68,546	34.77	24.49	24.25	16.50	40.74
	Big 4 Cities	7,081	62.75	19.31	12.94	5.01	17.95
	Urban/Suburban	12,746	50.29	25.35	18.54	5.82	24.36
	Rural	9,019	37.58	28.83	25.72	7.87	33.60
	Average Needs	36,846	25.45	26.61	32.45	15.49	47.94
	Low Needs	17,394	10.86	20.99	39.11	29.04	68.15
	Charter	10,049	25.52	27.31	30.28	16.89	47.17
	Religious and Independent	9,650	38.70	25.91	24.77	10.62	35.39
SWD	All Codes	27,573	67.56	19.82	10.06	2.57	12.62
SUA	All Codes	13,746	67.17	19.63	10.85	2.34	13.20
ELL	ELL=Y	16,297	68.55	19.94	8.89	2.63	11.51
SWD/ SUA	SWD & SUA codes	11,018	73.77	16.99	7.87	1.37	9.24
ELL/SUA	SUA & ELL codes	1,226	83.12	12.23	4.16	0.49	4.65

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
ELL Test Language	Chinese	686	14.87	21.57	38.05	25.51	63.56
	English	167,266	31.67	25.09	27.47	15.78	43.24
	Haitian-Creole	116	81.03	13.79	5.17		5.17
ELL Test Language	Korean	22	9.09	31.82	27.27	31.82	59.09
	Russian	119	36.97	26.05	26.89	10.08	36.97
	Spanish	3,234	76.78	16.94	5.50	0.77	6.28
	All Translations	171,443	32.49	24.91	27.08	15.52	42.60

8.2.2.4. Mathematics Grade 6

Table 8.28 presents the Mathematics Grade 6 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 39% of students achieved Level III and Level IV. About 41% of Female students were at Level III or above, as compared to 39% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (67%) students and students from Low Needs districts (67%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 17–25% of students in those same performance categories. Only about 10% of the SWD, SUA, and ELL subgroups, on average, earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (40%): Female (41%), Asian (67%), Multiracial (45%), Pacific Islander (45%), and White (50%) students, as well as those enrolled in Average (46%) and Low (67%) Needs districts and Charter schools (41%). For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 6% (Spanish) to 56% (Chinese).

Table 8.28. Mathematics Grade 6 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	167,034	29.96	30.68	18.78	20.58	39.36
Gender	Female	81,345	27.74	31.47	19.72	21.06	40.78
	Male	85,689	32.06	29.92	17.89	20.12	38.02
Ethnicity	Asian	18,038	11.89	20.81	21.23	46.07	67.30
	Black	31,860	46.03	31.96	12.99	9.02	22.01
	Hispanic	46,735	40.70	34.09	14.99	10.21	25.20
	American Indian	1,172	38.57	30.55	16.55	14.33	30.89
	Multiracial	3,116	26.48	28.21	19.83	25.48	45.31
	Pacific Islander	508	24.61	30.12	22.64	22.64	45.28
	White	65,605	19.52	30.46	23.59	26.43	50.02
NRC	New York	66,463	34.39	29.55	15.74	20.32	36.06
	Big 4 Cities	6,649	56.78	25.55	10.03	7.64	17.67
	Urban/Suburban	11,262	48.26	32.58	12.20	6.96	19.16

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
NRC	Rural	8,120	31.76	36.32	18.83	13.09	31.92
	Average Needs	34,087	21.37	32.36	23.64	22.63	46.28
	Low Needs	16,406	8.60	24.17	27.15	40.07	67.23
	Charter	10,836	25.53	33.43	20.64	20.39	41.04
	Religious and Independent	13,104	29.56	35.41	19.68	15.35	35.04
SWD	All Codes	26,644	67.04	23.96	6.07	2.93	9.00
SUA	All Codes	13,180	63.53	25.69	7.23	3.55	10.78
ELL	ELL=Y	14,865	66.15	23.83	6.25	3.77	10.02
SWD/ SUA	SWD & SUA codes	10,370	69.99	22.60	5.24	2.17	7.41
ELL/SUA	SUA & ELL codes	1,065	82.25	15.12	1.97	0.66	2.63
ELL Test Language	Chinese	839	15.14	29.20	23.84	31.82	55.66
	English	161,687	28.95	30.87	19.14	21.03	40.18
	Haitian-Creole	129	68.22	24.03	6.98	0.78	7.75
	Korean	35	25.71	22.86	25.71	25.71	51.43
	Russian	134	23.13	37.31	22.39	17.16	39.55
	Spanish	4,210	70.62	23.66	4.13	1.59	5.72
	All Translations	167,034	29.96	30.68	18.78	20.58	39.36

8.2.2.5. Mathematics Grade 7

Table 8.29 presents the Mathematics Grade 7 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 37% of students achieved Level III and Level IV. About 38% of Female students were at Level III or above, as compared to 37% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III and above were Asian (68%) students and students from Low Needs districts (64%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 12–23% of students in those same performance categories. Only about 9% of the SWD, SUA, and ELL subgroups, on average, earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (37%): Female (38%), Asian (68%), Multiracial (43%), Pacific Islander (43%), and White (47%) students, as well as those enrolled in Average (42%) and Low (64%) Needs districts and Charter schools (40%). For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 4% (Spanish) to 64% (Chinese).

Table 8.29. Mathematics Grade 7 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	155,255	32.98	29.62	24.01	13.40	37.40

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
Gender	Female	75,358	30.39	31.43	24.87	13.30	38.18
	Male	79,897	35.42	27.90	23.19	13.48	36.67
Ethnicity	Asian	17,576	12.17	20.11	30.52	37.20	67.72
	Black	29,690	49.69	29.83	15.44	5.04	20.48
	Hispanic	42,595	45.32	31.91	17.16	5.61	22.77
	American Indian	1,149	40.30	33.86	18.97	6.88	25.85
	Multiracial	2,429	29.07	27.91	24.17	18.86	43.02
	Pacific Islander	451	23.95	32.37	29.05	14.63	43.68
	White	61,365	22.37	30.61	31.09	15.93	47.01
NRC	New York	64,725	35.90	28.28	20.26	15.56	35.82
	Big 4 Cities	6,221	66.08	21.89	9.53	2.49	12.02
	Urban/Suburban	9,993	55.38	28.54	13.33	2.75	16.08
	Rural	7,699	37.62	35.26	21.55	5.57	27.12
	Average Needs	30,298	25.37	33.00	29.78	11.85	41.63
	Low Needs	16,467	11.16	25.07	38.85	24.92	63.76
	Charter	10,180	27.74	32.19	27.27	12.80	40.07
	Religious and Independent	9,545	31.40	34.72	24.74	9.15	33.88
SWD	All Codes	24,834	71.94	20.46	6.04	1.56	7.59
SUA	All Codes	11,803	68.45	22.11	7.52	1.92	9.44
ELL	ELL=Y	13,913	71.18	20.25	6.20	2.36	8.57
SWD/ SUA	SWD & SUA codes	9,455	75.19	18.80	4.94	1.07	6.01
ELL/SUA	SUA & ELL codes	888	86.37	12.16	1.46	0.00	1.46
ELL Test Language	Chinese	865	12.49	23.47	36.18	27.86	64.05
	English	150,075	31.89	29.95	24.49	13.67	38.16
	Haitian-Creole	140	67.14	25.00	6.43	1.43	7.86
	Korean	39	17.95	25.64	35.90	20.51	56.41
	Russian	147	38.78	34.69	18.37	8.16	26.53
	Spanish	3,989	77.14	18.43	4.04	0.40	4.44
	All Translations	155,255	32.98	29.62	24.01	13.40	37.40

8.2.2.6. Mathematics Grade 8

Table 8.30 presents the Mathematics Grade 8 performance level summaries and n-counts of demographic subgroups. Statewide, a combined 22% of students achieved Level III and Level IV. About 24% of Female students were at Level III or above, as compared to 20% of Male students. The percentage of students in Levels III and IV varied widely by ethnicity and NRC subgroup. The ethnicity and NRC category with the greatest percentages of students at Level III

and above were Asian (52%) students and students from Low Needs districts (37%). The Big 4 Cities, High Needs/Urban/Suburban, Black, and Hispanic students had a range of 5–15% of students in those same performance categories. Only about 5% of the SWD, SUA, and ELL subgroups, on average, earned at least a Level III. Each of the following subgroups had a higher percentage of students in Levels III and IV than statewide (22%): Female (24%), Asian (52%), Multiracial (23%), Pacific Islander (29%), and White (25%) students, as well as those enrolled in New York City (24%) and Low Needs districts (37%) and Charter (34%) and Religious and Independent (25%) schools. For ELL students who used translated test forms, the percentages of students earning at least a Level III ranged from 2% (Haitian-Creole) to 75% (Korean).

Table 8.30. Mathematics Grade 8 Performance Level Distribution by Subgroup

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
State	All Students	116,822	40.82	37.24	15.70	6.24	21.94
Gender	Female	55,480	36.32	39.53	17.15	6.99	24.14
	Male	61,342	44.89	35.17	14.38	5.56	19.94
Ethnicity	Asian	11,441	17.42	30.27	27.39	24.92	52.31
	Black	25,291	54.09	33.08	10.19	2.64	12.83
	Hispanic	35,720	48.29	36.27	12.09	3.35	15.44
	American Indian	948	50.21	34.70	11.08	4.01	15.08
	Multiracial	1,497	41.28	35.47	15.56	7.68	23.25
	Pacific Islander	354	35.88	35.59	19.49	9.04	28.53
	White	41,571	32.59	42.67	19.00	5.74	24.75
NRC	New York	53,231	41.02	34.78	15.57	8.62	24.19
	Big 4 Cities	5,546	69.92	22.47	6.17	1.44	7.61
	Urban/Suburban	7,595	64.45	30.03	4.86	0.66	5.52
	Rural	5,964	47.33	41.05	9.96	1.66	11.62
	Average Needs	18,802	37.43	45.70	15.01	1.85	16.87
	Low Needs	7,875	18.87	43.90	28.33	8.90	37.23
	Charter	6,573	27.66	37.91	24.17	10.25	34.43
	Religious and Independent	11,163	34.44	40.06	18.83	6.67	25.50
SWD	All Codes	21,096	73.24	22.57	3.57	0.62	4.20
SUA	All Codes	9,780	71.92	23.08	4.31	0.69	5.00
ELL	ELL=Y	12,327	68.22	24.47	5.45	1.86	7.31
SWD/ SUA	SWD & SUA codes	7,776	77.28	19.35	2.91	0.46	3.37
ELL/SUA	SUA & ELL codes	641	87.05	11.54	1.40	0.00	1.40
ELL Test Language	Chinese	755	16.03	27.81	29.54	26.62	56.16
	English	111,967	39.85	37.79	16.06	6.30	22.36
	Haitian-Creole	129	64.34	33.33	1.55	0.78	2.33

Demographic Category		N-Count	Performance Levels				
			Level I	Level II	Level III	Level IV	Level III & IV
	Korean	20	15.00	10.00	50.00	25.00	75.00
	Russian	139	39.57	36.69	16.55	7.19	23.74
	Spanish	3,812	73.69	23.32	2.54	0.45	2.99
	All Translations	116,822	40.82	37.24	15.70	6.24	21.94

Section 9: References

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.
- Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika* 37:29–51.
- Bock, R.D. & M. Aitkin (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika* 46:443–459.
- Cattell, R.B. (1966). The Screen Test for the Number of Factors. *Multivariate Behavioral Research* 1:245–276.
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16:297–334.
- Dorans, N.J., A.P. Schmitt & C.A. Bleistein (1992). The standardization approach to assessing comprehensive differential item functioning. *Journal of Educational Measurement* 29:309–319.
- Dorans, N.J. & P. W. Holland (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35–66). Hillsdale, NJ: Lawrence Erlbaum.
- Fleiss J.L. & J. Cohen (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and Psychological Measurement*, 33: 613–619.
- Green, D.R., W.M. Yen & G.R. Burket (1989). Experiences in the application of item response theory in test construction. *Applied Measurement in Education* 2:297–312.
- Huynh, H. & C. Schneider (2004). *Vertically moderated standards as an alternative to vertical scaling: assumptions, practices, and an odyssey through NAEP*. Paper presented at the National Conference on Large-Scale Assessment. Boston, MA, June 21.
- Jensen, A.R. (1980). *Bias in mental testing*. New York: Free Press.
- Johnson, N.L. & S. Kotz (1970). *Distributions in Statistics: Continuous Univariate Distributions*, Vol. 2. New York: John Wiley.
- Kim, S. & M. J. Kolen (2004). *STUIRT: A computer program for scale transformation under unidimensional item response theory models*. Iowa City, IA: Iowa Testing Programs, The University of Iowa.
- Kolen, M.J. & Z. Cui (2004). *POLYEQUATE*. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.

- Kolen, M.J. & R.L. Brennan (1995). *Test Equating: Methods and Practices*. New York: Springer-Verlag.
- Landis, J. R. & G. G. Koch. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159-174.
- Lee, W. C., B.A. Hanson & R.L. Brennan (2002). Estimating consistency and accuracy indices for multiple classifications. *Applied Psychological Measurement* 26:412–432.
- Lee, W. C. (2008). *Classification consistency and accuracy for complex assessments using item response theory*. (CASMA Research Report No. 27). Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.
- Lee, W. C. & M. J. Kolen (2006, Revised 2008). IRT-CLASS (Version 2.0). Iowa City, IA: Center for Advanced Studies in Measurement and Assessment, The University of Iowa.
- Linn, R.L. (1991). Linking results of distinct assessments. *Applied Measurement in Education* 6(1): 83–102.
- Linn, R.L. & D. Harnisch (1981). Interactions between item content and group membership on achievement test items. *Journal of Educational Measurement* 18: 109–118.
- Livingston, S.A. & C. Lewis (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement* 32: 179–197.
- Lord, F.M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ: Lawrence Erlbaum.
- Lord, F.M. & M.R. Novick (1968). *Statistical Theories of Mental Test Scores*. Menlo Park, CA: Addison-Wesley.
- Mehrens, W.A. & I.J. Lehmann (1991). *Measurement and Evaluation in Education and Psychology*, 3rd ed. New York: Holt, Rinehart, and Winston.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement* 16: 159–176.
- Muraki, E. & R.D. Bock (1991). *PARSCALE: Parameter Scaling of Rating Data* [Computer program]. Chicago, IL: Scientific Software, Inc.
- Novick, M.R. & P.H. Jackson (1974). *Statistical Methods for Educational and Psychological Research*. New York: McGraw-Hill.
- NYSED. (2013) New York State Testing Program 2013: English Language Arts and Mathematics Grades 3–8 Technical Report. Albany, NY: New York State Education Department (NYSED). Retrieved from: <http://www.p12.nysed.gov/assessment/reports/2013/ela-math-tr13.pdf>
- Qualls, A.L. (1995). Estimating the reliability of a test containing multiple-item formats. *Applied Measurement in Education* 8: 111–120.

- Reckase, M.D. (1979). Unifactor latent trait models applied to multifactor tests: results and implications. *Journal of Educational Statistics* 4: 207–230.
- Sandoval, J.H. & M.P. Mille (1979) *Accuracy of judgments of WISC-R item difficulty for minority groups*. Paper presented at the annual meeting of the American Psychological Association, New York. August.
- Stocking, M.L. & F.M. Lord (1983). Developing a common metric in item response theory. *Applied Psychological Measurement* 7: 201–210.
- Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika* 47: 175–186.
- Cai, L., Thissen, D. J., & du Toit, S. (2011). IRTPRO (Version 2.1). Skokie, IL: Scientific Software International, Inc.
- Thompson, S.J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal Design Applied to Large Scale Assessments (NCEO Synthesis Report 44)*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved from: <http://www.cehd.umn.edu/nceo/onlinepubs/Synthesis44.html>.
- Wang, T.M., J. Kolen, & D.J. Harris (2000). Psychometric properties of scale scores and performance levels for performance assessment using polytomous IRT. *Journal of Educational Measurement* 37: 141–162.
- Yen, W.M. (1997). The technical quality of performance assessments: Standard errors of percents of students reaching standards. *Educational Measurement: Issues and Practice*: 5–15.
- Yen, W.M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement* 30: 187–213.
- Yen, W. M. (1984). Obtaining maximum likelihood trait estimates from number correct scores for the three-parameter logistic model. *Journal of Educational Measurement* 21:93–111.
- Yen, W.M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement* 5: 245–262.
- Yen, W.M., R.C. Sykes, K. Ito & M. Julian (1997). *A Bayesian/IRT index of objective performance for tests with mixed-item types*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago: March.
- Zwick, R., J.R. Donoghue & A. Grima, (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement* 36: 225–33.

Appendix A: ELA and Mathematics Test Configurations

Table A1. ELA Test Configuration

Grade	Day Book		Number of Items				
			Multiple-Choice		Constructed-Response		Total
			Operational	Embedded	Operational	Embedded	
3	1	1	18	6	0	0	24
	2	2	7	0	3	0	10
	3	3	0	0	6	0	6
	Total		25	6	9	0	40
4	1	1	18	6	0	0	24
	2	2	7	0	3	0	10
	3	3	0	0	6	0	6
	Total		25	6	9	0	40
5	1	1	28	7	0	0	35
	2	2	7	0	3	0	10
	3	3	0	0	6	0	6
	Total		35	7	9	0	51
6	1	1	28	7	0	0	35
	2	2	7	0	3	0	10
	3	3	0	0	6	0	6
	Total		35	7	9	0	51
7	1	1	28	7	0	0	35
	2	2	7	0	3	0	10
	3	3	0	0	6	0	6
	Total		35	7	9	0	51
8	1	1	28	7	0	0	35
	2	2	7	0	3	0	10
	3	3	0	0	6	0	6
	Total		35	7	9	0	51

Table A2. Mathematics Test Configuration

Grade	Day Book		Number of Items				
			Multiple-Choice		Constructed-Response		Total
			Operational	Embedded	Operational	Embedded	
3	1	1	18	4	0	0	22
	2	2	19	3	0	0	22
	3	3	0	0	8	0	8
	Total		37	7	8	0	52
4	1	1	18	4	0	0	22
	2	2	20	3	0	0	23
	3	3	0	0	10	0	10
	Total		38	7	10	0	55
5	1	1	18	4	0	0	22
	2	2	20	3	0	0	23
	3	3	0	0	10	0	10
	Total		38	7	10	0	55
6	1	1	22	4	0	0	26
	2	2	22	3	0	0	25
	3	3	0	0	10	0	10

Appendix A: ELA and Mathematics Test Configurations and Testing Times

Grade	Day Book		Number of Items				
			Multiple-Choice		Constructed-Response		Total
			Operational	Embedded	Operational	Embedded	
	Total		44	7	10	0	61
7	1	1	22	4	0	0	26
	2	2	22	3	0	0	25
	3	3	0	0	10	0	10
	Total		44	7	10	0	61
8	1	1	22	4	0	0	26
	2	2	22	3	0	0	25
	3	3	0	0	10	0	10
	Total		44	7	10	0	61

Table A3. ELA Estimated Time on Task by Book

Grade	Day	Book	Estimated Time on Task (min.)	Previous Session Time (min.)
3	1	1	50	70
	2	2	50	70
	3	3	50	70
		Total	150	210
4	1	1	50	70
	2	2	50	70
	3	3	50	70
		Total	150	210
5	1	1	60	90
	2	2	50	90
	3	3	50	90
		Total	160	270
6	1	1	60	90
	2	2	50	90
	3	3	50	90
		Total	160	270
7	1	1	60	90
	2	2	50	90
	3	3	50	90
		Total	160	270
8	1	1	60	90
	2	2	50	90
	3	3	50	90
		Total	160	270

Source: 2017 ELA and Mathematics Test Guides.

The ELA estimated times on task were based on the following rules of thumb:

- Average time to read a passage—5 minutes
- Average time to respond to a multiple-choice question—1 minute
- Average time to respond to a two-point constructed response question—3 minutes
- Average time to respond to a four-point constructed response question—20 minutes

Table A4. Mathematics Estimated Time on Task by Book

Grade	Day	Book	Estimated Time on Task (min.)	Previous Session Time (min.)
3	1	1	40	60
	2	2	40	60
	3	3	60	70
	Total		140	190
4	1	1	40	60
	2	2	40	60
	3	3	70	90
	Total		150	210
5	1	1	40	80
	2	2	40	80
	3	3	70	90
	Total		150	250
6	1	1	40	80
	2	2	40	80
	3	3	70	90
	Total		150	250
7	1	1	40	80
	2	2	40	80
	3	3	70	90
	Total		150	250
8	1	1	40	80
	2	2	40	80
	3	3	70	90
	Total		150	250

Source: 2017 ELA and Mathematics Test Guides.

The Mathematics estimated times on task were based on the following rules of thumb:

- Average time to respond to a multiple-choice question—1.5 minutes
- Average time to respond to a two-point constructed response question—5 minutes
- Average time to respond to a three-point constructed response question—9 minutes

The testing times listed above do not include approximately 10 minutes reserved for preparation at the beginning of each session for handing out materials and reading directions. Additional details on security, scheduling, classroom organization and preparation, test materials, and administration can be found in the *2017 Teacher's Directions* and the *School Administrator's Manual*, which are accessible online:

- *2017 ELA Teacher's Directions*
 - Grades 3–5: <http://www.p12.nysed.gov/assessment/ei/2017/td-ela-g35-17.pdf>
 - Grades 6–8: <http://www.p12.nysed.gov/assessment/ei/2017/td-ela-g68-17.pdf>
- *2017 Mathematics Teacher's Directions*
 - Grades 3–5: <http://www.p12.nysed.gov/assessment/ei/2017/td-math-g35-17.pdf>

Appendix A: ELA and Mathematics Test Configurations and Testing Times

- Grades 6–8: <http://www.p12.nysed.gov/assessment/ei/2017/td-math-g68-17.pdf>
- *2017 ELA and Mathematics Tests School Administrator’s Manual*
 - Volume 1: <http://www.p12.nysed.gov/assessment/sam/ei/eisam17-v1.pdf>
 - Volume 2: <http://www.p12.nysed.gov/assessment/sam/ei/eisam17-v2.pdf>
- *2017 ELA and Mathematics Test Guides*
 - <https://www.engageny.org/resource/test-guides-english-language-arts-and-mathematics>

Appendix B: ELA and Mathematics Test Blueprints

Table B1. ELA Test Blueprint

Grade	Total Points on OP Test	Strand	Point Range		% of Test	
			Target	Actual	Target	Actual
3	47	Literature	14–44	22	30%-94%	47%
		Information	14–44	24	30%-94%	51%
		Language	1–4	1	2%-9%	2%
4	47	Literature	14–44	23	30%-94%	49%
		Information	14–44	23	30%-94%	49%
		Language	1–4	1	2%-9%	2%
5	57	Literature	18–51	22	32%-89%	39%
		Information	18–51	33	32%-89%	58%
		Language	1–4	2	2%-7%	4%
6	57	Literature	11–44	24	19%-77%	42%
		Information	25–58	32	44%-102%	56%
		Language	1–4	1	2%-7%	2%
7	57	Literature	11–44	21	19%-77%	37%
		Information	25–58	34	44%-102%	60%
		Language	1–4	2	2%-7%	4%
8	57	Literature	11–44	23	19%-77%	40%
		Information	25–58	33	44%-102%	58%
		Language	1–4	1	2%-7%	2%

Table B2. Mathematics Test Blueprint

Grade	Total Points on OP Test	Domain	Point Range		% of Test	
			Target	Actual	Target	Actual
3	56	Operations and Algebraic Thinking	23–31	25	41%-55%	45%
		Number and Operations in Base Ten	3–5	4	5%-9%	7%
		Number and Operations – Fractions	10–14	11	18%-25%	20%
		Measurement and Data	12–18	13	21%-32%	23%
		Geometry*	1–3	4	2%-5%	7%
4	62	Operations and Algebraic Thinking	11–15	11	18%-24%	18%
		Number and Operations in Base Ten	14–20	16	23%-32%	26%
		Number and Operations – Fractions	15–21	17	24%-34%	27%

Appendix B: ELA and Mathematics Test Blueprints

Grade	Total Points on OP Test	Domain	Point Range		% of Test	
			Target	Actual	Target	Actual
4	62	Measurement and Data	9–15	12	15%-24%	19%
		Geometry	5–7	5	8%-11%	9%
5	62	Operations and Algebraic Thinking	3–5	4	5%-8%	7%
		Number and Operations in Base Ten	15–21	16	24%-34%	26%
		Number and Operations – Fractions	22–28	23	35%-45%	37%
		Measurement and Data	12–18	15	19%-29%	24%
		Geometry*	1–3	5	2%-5%	9%
6	68	Ratios and Proportional Relationships	16–20	17	24%-29%	25%
		The Number System	13–19	15	19%-28%	22%
		Expressions and Equations	23–33	26	34%-49%	38%
		Geometry	8–12	10	14%-21%	15%
7	68	Ratios and Proportional Relationships	18–22	18	26%-32%	26%
		The Number System	12–16	15	18%-24%	22%
		Expressions and Equations	19–25	20	28%-37%	29%
		Geometry	3–7	8	5%-13%	14%
		Statistics and Probability	8–14	8	14%-25%	12%
8	68	Expressions and Equations	26–34	29	38%-50%	43%
		Functions	16–22	18	24%-32%	26%
		Geometry	14–20	14	21%-29%	21%
		Statistics and Probability	5–7	5	9%-13%	9%

*There is a slight difference between the “Target% of Test” shown in these tables and the tables presented in the Guides to the 2017 Mathematics Tests. The guides were intended to provide general guidance regarding content coverage of mathematics domains so that classroom instruction would continue to cover the depth and breadth of the mathematics standards.

Appendix C: Passage Selection Guidelines for Assessing ELA

General Guidelines

Along with instructional materials and teacher training, assessment development is essential to the successful implementation of the CCSS. While many of the expectations outlined in the CCSS align with previous versions of the New York State Learning Standards for ELA, the CCSS do represent some shifts in emphasis with direct implications for assessment development. In particular, the CCSS devote considerable attention to the types and nature of texts used in instruction and assessment. The foundation for preparing students for the linguistic rigors of college and of the workplace lies in the texts with which they interact. By the time that they graduate, students should be prepared to successfully read and analyze the types of complex texts that they will encounter after high school. Selecting passages of appropriate type and complexity for use in assessment is integral to this preparation.

One of the major shifts of the CCSS is an emphasis on developing skills for comprehending and analyzing informational texts. Increased exposure to informational texts better prepares students for the various types of texts that they will encounter in college and in the workplace. The array of passages selected for assessment from K–12 should support the development of the necessary skills to handle this range of informational texts.

Another shift is an increased emphasis on the analysis across multiple texts, often of varied genres and media. Several standards, especially for reading literature, require intertextual and multi-media analysis. These expectations require special attention to the selection of related passages, chosen specifically to support the assessment of the full range of expectations. It will also require careful consideration of which standards are appropriate for large-scale assessment formats, and how these assessments might be modified to include passages of a variety of media.

In addition to the usual fairness and sensitivity guidelines when selecting passages for assessment, attention should be dedicated to three additional considerations:

- *Text Complexity*
- *Text Types*
- *Text Suitability for Specific Standards*

These guidelines should inform the training of passage finders, in order to ensure a pool of acceptable passages that can support assessment of all the CCSS Reading Informational Texts standards. They should also alert form assemblers as they construct forms that will assess the complete range of skills.

Appendix D: Universal Design Item Checklist

Universal Design Item Checklist	
A.	Precisely Designed Constructs
Definition	The item construct is clearly defined so that all irrelevant cognitive, sensory, emotional, and physical barriers are removed.
√	The item does not add skills to those being measured (no extraneous skills tested).
B.	Language Appropriateness
Definition	The item avoids words or phrases that are sexist, racist, or otherwise offensive, inappropriate, or negative to any subgroup. Language should be simple and clear.
√	The item uses commonly used words—simpler is better.
√	The item uses vocabulary appropriate for the grade level.
√	Idiomatic speech and figurative language are avoided unless being measured.
√	The item avoids technical terms unrelated to the content.
√	The item contains no unnecessary words.
√	The sentence complexity contained in the item is appropriate for the grade level.
√	The item avoids ambiguous or multiple-meaning words (e.g., crane—the bird—can easily be confused with crane—heavy machinery).
√	All pronouns have clear referents.
√	The item avoids the use of proper names. (Such names may be unfamiliar or difficult for cultural subgroups.)
√	The item avoids irregularly spelled words.
C.	Gender Stereotypes
Definition	The item avoids stereotyping as results of associating genders with certain professions or activities. All groups of society should be portrayed accurately and fairly regarding gender.
√	The item is free of content that might offend a gender subgroup.
√	The item is free of content that might unfairly advantage or disadvantage a gender subgroup.
D.	Ethnic Stereotypes
Definition	The item avoids unnecessary references to and uses the proper reference for ethnic, racial, or cultural groups.
√	The item is free of content that might offend an ethnic subgroup.
√	The item is free of content that might unfairly advantage or disadvantage an ethnic subgroup.
√	The artwork included in an item adequately reflects the diversity of the student population.
E.	Cultural Familiarity
Definition	Does not rely on an assumed shared experience that is class oriented or native English speaking oriented. Presentations of cultural or ethnic differences should neither explicitly nor implicitly rely on stereotypes nor make moral judgments.
√	The item does not rely on an assumed shared experience that is class oriented or native English speaking oriented.
√	The item is free from content that might offend a socioeconomic subgroup.
√	The item is free of content that might unfairly advantage or disadvantage a socioeconomic subgroup.

Appendix D: Universal Design Item Checklist

Universal Design Item Checklist	
√	The item is free from unnecessary cultural references.
√	The item is free from religious references.
F.	Geographic Bias
Definition	All groups of society should be portrayed accurately and fairly regarding geographic setting. A particular geographic setting shouldn't be used repeatedly, and urban, suburban, and rural settings should be represented across items.
√	The item is free of content that might offend a geographic subgroup.
√	The item is free of content that might unfairly advantage or disadvantage a geographic subgroup.
G.	Disability Bias
Definition	All groups of society should be portrayed accurately and fairly regarding disability. Stereotypes related to any particular disability should be avoided. No undue restrictions should exist in the item that would interfere with the ability of a student to comprehend or respond to the item.
√	The item is free of content that might offend a disability subgroup.
√	The item is free of content that might unfairly advantage or disadvantage a disability subgroup.
√	A graphic representation is used in the items, as appropriate. The complexity of the graphic is appropriate to the purpose—simpler is better.
√	The item avoids content that depends on sensory knowledge (such as references to movement, sound, smell, etc.) unless this is crucial to the overall item.
√	The item could be put into Braille.
√	The item avoids using both O and Q.
√	Letter pairs can be easily distinguished when read. (S and T are okay; S and X are not).
H.	Art Supports Text
Definition	The art is related to the item and supports the reader when possible. The item text and art are legible and accessible, and the art is appropriately placed in the item to support the reader. The art does not distract the test taker, but instead provides a scaffold to overall comprehension.
√	All pictures relate to items.
√	The item is free from pictorial clutter: All pictures are needed to answer the item.
√	Graphics are clear and non-fuzzy.
√	Any symbols used are highly distinguishable.
√	Visual load requirements are reasonable for the grade level.
√	Multi-dimensional graphics and complex shading are avoided.
√	Tables have replaced any cluttered graphs.
√	Labels read clockwise (as is easier for Braille readers).
I.	Special Populations Considerations
Definition	Consideration must be given for maximum accessibility to all students including, but not limited to, English language learners, limited sight, hearing impaired, cognitively challenged, etc. These considerations will assist all students.
√	The item contains scaffolding techniques to support student understanding of what is being asked in the item.
√	Text is replaced with graphic representations, when appropriate.
√	The item is written with simplified text load.
√	The item is written with simplified sentences.

Appendix D: Universal Design Item Checklist

Universal Design Item Checklist	
√	The item has as little extraneous information as possible.
√	The item provides context, but it is simplified.
√	The item uses smaller or less complicated numbers or expressions where not otherwise required.
√	The item avoids negative phrasing or questions; for example, questions are not asked in the negative.

Appendix E: Criteria for Item Acceptability

The following criteria represent best practices in item development, and were implemented during the creation and review of the New York State 3–8 CCSS test questions; however, these criteria are not a substitute for the full, detailed criteria documents, which are available online at the following links:

- <http://www.engageny.org/resource/new-york-state-item-review-criteria-for-grade-3-8-english-language-arts-tests>; and
- <http://www.engageny.org/resource/new-york-state-item-review-criteria-for-grade-3-8-mathematics-tests>.

For Multiple-Choice Items:

Check that the content of each item:

- is targeted to assess only one objective or skill (unless specifications indicate otherwise)
- deals with material that is important in testing the targeted performance indicator
- uses grade-appropriate content and thinking skills
- is presented at a reading level suitable for the grade level being tested
- has a stem that facilitates answering the question or completing the statement without looking at the answer choices
- has a stem that does **not** present clues to the correct answer choice
- has answer choices that are plausible and attractive to the student who has not mastered the objective or skill
- has mutually exclusive distractors
- has one and only one correct answer choice
- is free of cultural, racial, ethnic, age, gender, disability, regional, or other apparent bias

Check that the format of each item:

- is worded in the positive unless it is absolutely necessary to use the negative form
- is free of extraneous words or expressions in both the stem and the answer choices (e.g., the same word or phrase does not begin each answer choice)
- indicates emphasis on key words, such as best, first, least, not, and others that are important and might be overlooked
- places the interrogative word at the **beginning** of a stem in the form of a question, or places the omitted portion of an incomplete statement at the **end** of the statement
- indicates the correct answer choice
- provides the rationale for all distractors
- is conceptually, grammatically, and syntactically consistent—between the stem and answer choices, and among the answer choices
- has answer choices balanced in length, or contains two long and two short answer choices
- clearly identifies the passage or other stimulus material associated with the item
- clearly identifies a need of for art, if applicable, and the art is conceptualized and sketched, with important considerations explicated

Also check that:

- one item does not present clues to the correct answer choice for any other item
- any item based on a passage is answerable from the information given in the passage and is not dependent on skills related to other content areas
- any item based on a passage is truly passage-dependent; that is, **not** answerable without reference to the passage
- there is a balance of reasonable, non-stereotypical representation of economic classes, races, cultures, ages, genders, and persons with disabilities in context and art

For Constructed-Response Items:

Check that the content of each item is:

- designed to assess the targeted performance indicator
- appropriate for the grade level being tested
- presented at a reading level suitable for the grade level being tested
- appropriate in context
- written so that a student possessing knowledge or skill being tested can construct a response that can be scored with the specified rubric or scoring tool; that is, the range of possible correct responses must be wide enough to allow for a diversity of responses, but narrow enough so that students who do not clearly show their grasp of the objective or skill being assessed cannot obtain the maximum score
- presented without clues to the correct response
- checked for accuracy and documented against reliable, up-to-date sources (including rubrics)
- free of cultural, racial, ethnic, age, gender, disability, or other apparent bias

Check that the format of each item is:

- appropriate for the question being asked and the intended response
- worded clearly and concisely, using simple vocabulary and sentence structure
- precise and unambiguous in its directions for the desired response
- free of extraneous words or expressions
- worded in the positive form rather than in the negative form
- conceptually, grammatically, and syntactically consistent
- marked with emphasis on key words, such as best, first, least, and others that are important and might be overlooked
- clearly identified as needing art, if applicable, and the art is conceptualized and sketched, with important considerations explicated

Also check that:

- one item does not present clues to the correct response to any other item
- there is a balance of reasonable, non-stereotypical representation of economic classes, races, cultures, ages, genders, and persons with disabilities in context and art
- for each set of items related to a reading passage, each item is designed to elicit a unique and independent response
- items designed to assess reading do not depend on prior knowledge of the subject matter used in the prompt/question

Appendix F: Psychometric Guidelines for Operational Item Selection

It is primarily up to the content development department to select items for the 2017 Operational Test. The psychometrics department will provide support, as necessary, and will review the final item selection. The psychometrics department will provide data files with parameters for all FT items eligible for the item pool. The pools of items eligible for 2017 item selection included 2013, 2014, 2015, 2016 embedded and stand-alone field-test items.

Here are the general guidelines for item selection:

- Satisfy the content specifications in terms of objective coverage and the number and percentage of MC and CR items on the test. An often-used criterion for objective coverage is within 5% of the percentages of score points and items per objective.
- To the extent possible, select both easy and difficult items to provide good measurement information at both ends of the performance scale.
- Avoid selecting items with too high/low p -values, items with flagged point biserials, and poorly fitting items.
- Minimize the number of items flagged for DIF (gender, ethnic, and High/Low Needs schools). Flagged items should be reviewed for content again. It needs to be remembered that some items may be flagged for DIF by chance only, and that their content may not necessarily be biased against any of the analyzed subgroups. The psychometrics department will provide DIF information for each item. It is also possible to get “significant” DIF, but not bias, if the content is a necessary part of the construct that is measured. That is, there may be some non-false positive DIF flags on items that do not exhibit bias.
- Provide the NYSED with the following summary information:
 - Overview of the statistical properties of the tests
 - Blueprint comparison between the test build and the target. The focus is on the total number of points on the test
 - Raw score proportion correct comparison between the test build and the reference (i.e., Spring 2016 test)
 - Vertically linked average difficulty parameter (MC items only) across all grades
 - Vertically linked TCC based on the constructed test
 - TCC, Test Information Curves and Conditional SEM Curves for each subject and grade, again using the Spring 2016 operational test as a reference.

Appendix G: Operational Item Maps

The following tables show the operational item maps for the 2017 NYSTP Grades 3–8 ELA and Mathematics Tests. External linking and field test items (i.e., those not contributing to students' scores) have been omitted. Additional detail on the standards to which these items align may be found at: <http://www.engageny.org/resource/new-york-state-p-12-common-core-learning-standards>.

Table G1. ELA Grade 3 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RL.3.3
2	MC	1	CCSS.ELA-Literacy.RL.3.4
3	MC	1	CCSS.ELA-Literacy.RL.3.5
4	MC	1	CCSS.ELA-Literacy.RL.3.3
5	MC	1	CCSS.ELA-Literacy.RL.3.2
6	MC	1	CCSS.ELA-Literacy.RL.3.3
7	MC	1	CCSS.ELA-Literacy.L.3.5a
8	MC	1	CCSS.ELA-Literacy.RI.3.8
9	MC	1	CCSS.ELA-Literacy.RI.3.1
10	MC	1	CCSS.ELA-Literacy.RI.3.5
11	MC	1	CCSS.ELA-Literacy.RI.3.2
12	MC	1	CCSS.ELA-Literacy.RI.3.1
19	MC	1	CCSS.ELA-Literacy.RL.3.1
20	MC	1	CCSS.ELA-Literacy.RL.3.5
21	MC	1	CCSS.ELA-Literacy.RL.3.1
22	MC	1	CCSS.ELA-Literacy.RL.3.7
23	MC	1	CCSS.ELA-Literacy.RL.3.3
24	MC	1	CCSS.ELA-Literacy.RL.3.2
25	MC	1	CCSS.ELA-Literacy.RI.3.3
26	MC	1	CCSS.ELA-Literacy.RI.3.2
27	MC	1	CCSS.ELA-Literacy.RI.3.1
28	MC	1	CCSS.ELA-Literacy.RI.3.8
29	MC	1	CCSS.ELA-Literacy.RI.3.4
30	MC	1	CCSS.ELA-Literacy.RI.3.3
31	MC	1	CCSS.ELA-Literacy.RI.3.1
32	CR	2	CCSS.ELA-Literacy.RL.3.4
33	CR	2	CCSS.ELA-Literacy.RL.3.3
34	CR	4	CCSS.ELA-Literacy.RI.3.3
35	CR	2	CCSS.ELA-Literacy.RI.3.2
36	CR	2	CCSS.ELA-Literacy.RI.3.5
37	CR	2	CCSS.ELA-Literacy.RI.3.8

Item	Type	Points	Standard
38	CR	2	CCSS.ELA-Literacy.RI.3.7
39	CR	2	CCSS.ELA-Literacy.RL.3.5
40	CR	4	CCSS.ELA-Literacy.RL.3.2

Table G2. ELA Grade 4 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RI.4.1
2	MC	1	CCSS.ELA-Literacy.RI.4.1
3	MC	1	CCSS.ELA-Literacy.RI.4.1
4	MC	1	CCSS.ELA-Literacy.RI.4.3
5	MC	1	CCSS.ELA-Literacy.RI.4.5
6	MC	1	CCSS.ELA-Literacy.RI.4.2
7	MC	1	CCSS.ELA-Literacy.RL.4.4
8	MC	1	CCSS.ELA-Literacy.RL.4.1
9	MC	1	CCSS.ELA-Literacy.RL.4.6
10	MC	1	CCSS.ELA-Literacy.RL.4.3
11	MC	1	CCSS.ELA-Literacy.RL.4.3
12	MC	1	CCSS.ELA-Literacy.RL.4.2
19	MC	1	CCSS.ELA-Literacy.RL.4.1
20	MC	1	CCSS.ELA-Literacy.RL.4.3
21	MC	1	CCSS.ELA-Literacy.L.4.5a
22	MC	1	CCSS.ELA-Literacy.RL.4.3
23	MC	1	CCSS.ELA-Literacy.RL.4.2
24	MC	1	CCSS.ELA-Literacy.RL.4.2
25	MC	1	CCSS.ELA-Literacy.RI.4.1
26	MC	1	CCSS.ELA-Literacy.RI.4.3
27	MC	1	CCSS.ELA-Literacy.RI.4.7
28	MC	1	CCSS.ELA-Literacy.RI.4.8
29	MC	1	CCSS.ELA-Literacy.RI.4.4
30	MC	1	CCSS.ELA-Literacy.RI.4.2
31	MC	1	CCSS.ELA-Literacy.RI.4.1
32	CR	2	CCSS.ELA-Literacy.RL.4.4
33	CR	2	CCSS.ELA-Literacy.RL.4.3
34	CR	4	CCSS.ELA-Literacy.RL.4.2
35	CR	2	CCSS.ELA-Literacy.RL.4.4
36	CR	2	CCSS.ELA-Literacy.RL.4.6
37	CR	2	CCSS.ELA-Literacy.RI.4.4
38	CR	2	CCSS.ELA-Literacy.RI.4.1

Item	Type	Points	Standard
39	CR	2	CCSS.ELA-Literacy.RI.4.8
40	CR	4	CCSS.ELA-Literacy.RI.4.9

Table G3. ELA Grade 5 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RL.5.3
2	MC	1	CCSS.ELA-Literacy.RL.5.4
3	MC	1	CCSS.ELA-Literacy.RL.5.1
4	MC	1	CCSS.ELA-Literacy.RL.5.1
5	MC	1	CCSS.ELA-Literacy.RL.5.7
6	MC	1	CCSS.ELA-Literacy.RL.5.5
7	MC	1	CCSS.ELA-Literacy.RL.5.2
8	MC	1	CCSS.ELA-Literacy.RI.5.2
9	MC	1	CCSS.ELA-Literacy.RI.5.1
10	MC	1	CCSS.ELA-Literacy.RI.5.4
11	MC	1	CCSS.ELA-Literacy.RI.5.1
12	MC	1	CCSS.ELA-Literacy.RI.5.3
13	MC	1	CCSS.ELA-Literacy.RI.5.3
14	MC	1	CCSS.ELA-Literacy.RI.5.2
22	MC	1	CCSS.ELA-Literacy.RL.5.4
23	MC	1	CCSS.ELA-Literacy.RL.5.5
24	MC	1	CCSS.ELA-Literacy.RL.5.1
25	MC	1	CCSS.ELA-Literacy.RL.5.2
26	MC	1	CCSS.ELA-Literacy.RL.5.3
27	MC	1	CCSS.ELA-Literacy.RL.5.6
28	MC	1	CCSS.ELA-Literacy.RL.5.2
29	MC	1	CCSS.ELA-Literacy.L.5.5b
30	MC	1	CCSS.ELA-Literacy.RI.5.1
31	MC	1	CCSS.ELA-Literacy.RI.5.8
32	MC	1	CCSS.ELA-Literacy.RI.5.8
33	MC	1	CCSS.ELA-Literacy.RI.5.1
34	MC	1	CCSS.ELA-Literacy.RI.5.5
35	MC	1	CCSS.ELA-Literacy.RI.5.3
36	MC	1	CCSS.ELA-Literacy.L.5.5a
37	MC	1	CCSS.ELA-Literacy.RI.5.5
38	MC	1	CCSS.ELA-Literacy.RI.5.8
39	MC	1	CCSS.ELA-Literacy.RI.5.3
40	MC	1	CCSS.ELA-Literacy.RI.5.1

Item	Type	Points	Standard
41	MC	1	CCSS.ELA-Literacy.RI.5.7
42	MC	1	CCSS.ELA-Literacy.RI.5.2
43	CR	2	CCSS.ELA-Literacy.RI.5.5
44	CR	2	CCSS.ELA-Literacy.RI.5.1
45	CR	4	CCSS.ELA-Literacy.RL.5.2
46	CR	2	CCSS.ELA-Literacy.RL.5.3
47	CR	2	CCSS.ELA-Literacy.RL.5.4
48	CR	2	CCSS.ELA-Literacy.RI.5.4
49	CR	2	CCSS.ELA-Literacy.RI.5.2
50	CR	2	CCSS.ELA-Literacy.RI.5.8
51	CR	4	CCSS.ELA-Literacy.RI.5.9

Table G4. ELA Grade 6 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RI.6.1
2	MC	1	CCSS.ELA-Literacy.RI.6.5
3	MC	1	CCSS.ELA-Literacy.RI.6.4
4	MC	1	CCSS.ELA-Literacy.RI.6.1
5	MC	1	CCSS.ELA-Literacy.RI.6.3
6	MC	1	CCSS.ELA-Literacy.RI.6.8
7	MC	1	CCSS.ELA-Literacy.RI.6.8
8	MC	1	CCSS.ELA-Literacy.L.6.5a
9	MC	1	CCSS.ELA-Literacy.RL.6.5
10	MC	1	CCSS.ELA-Literacy.RL.6.4
11	MC	1	CCSS.ELA-Literacy.RL.6.4
12	MC	1	CCSS.ELA-Literacy.RL.6.2
13	MC	1	CCSS.ELA-Literacy.RL.6.1
14	MC	1	CCSS.ELA-Literacy.RL.6.3
15	MC	1	CCSS.ELA-Literacy.RL.6.1
16	MC	1	CCSS.ELA-Literacy.RL.6.1
17	MC	1	CCSS.ELA-Literacy.RL.6.1
18	MC	1	CCSS.ELA-Literacy.RL.6.3
19	MC	1	CCSS.ELA-Literacy.RL.6.5
20	MC	1	CCSS.ELA-Literacy.RL.6.3
21	MC	1	CCSS.ELA-Literacy.RL.6.6
29	MC	1	CCSS.ELA-Literacy.RI.6.3
30	MC	1	CCSS.ELA-Literacy.RI.6.1
31	MC	1	CCSS.ELA-Literacy.RI.6.3

Item	Type	Points	Standard
32	MC	1	CCSS.ELA-Literacy.RI.6.3
33	MC	1	CCSS.ELA-Literacy.RI.6.5
34	MC	1	CCSS.ELA-Literacy.RI.6.2
35	MC	1	CCSS.ELA-Literacy.RI.6.2
36	MC	1	CCSS.ELA-Literacy.RL.6.3
37	MC	1	CCSS.ELA-Literacy.RL.6.5
38	MC	1	CCSS.ELA-Literacy.RL.6.4
39	MC	1	CCSS.ELA-Literacy.RL.6.5
40	MC	1	CCSS.ELA-Literacy.RL.6.2
41	MC	1	CCSS.ELA-Literacy.RL.6.1
42	MC	1	CCSS.ELA-Literacy.RL.6.2
43	CR	2	CCSS.ELA-Literacy.RI.6.4
44	CR	2	CCSS.ELA-Literacy.RI.6.5
45	CR	4	CCSS.ELA-Literacy.RI.6.6
46	CR	2	CCSS.ELA-Literacy.RL.6.3
47	CR	2	CCSS.ELA-Literacy.RL.6.2
48	CR	2	CCSS.ELA-Literacy.RI.6.1
49	CR	2	CCSS.ELA-Literacy.RI.6.2
50	CR	2	CCSS.ELA-Literacy.RI.6.4
51	CR	4	CCSS.ELA-Literacy.RI.6.9

Table G5. ELA Grade 7 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RI.7.1
2	MC	1	CCSS.ELA-Literacy.RI.7.1
3	MC	1	CCSS.ELA-Literacy.RI.7.2
4	MC	1	CCSS.ELA-Literacy.RI.7.1
5	MC	1	CCSS.ELA-Literacy.RI.7.4
6	MC	1	CCSS.ELA-Literacy.RI.7.2
7	MC	1	CCSS.ELA-Literacy.RI.7.2
8	MC	1	CCSS.ELA-Literacy.RI.7.8
9	MC	1	CCSS.ELA-Literacy.RI.7.3
10	MC	1	CCSS.ELA-Literacy.RI.7.4
11	MC	1	CCSS.ELA-Literacy.RI.7.1
12	MC	1	CCSS.ELA-Literacy.RI.7.1
13	MC	1	CCSS.ELA-Literacy.RI.7.2
14	MC	1	CCSS.ELA-Literacy.RI.7.2
22	MC	1	CCSS.ELA-Literacy.RL.7.1

Item	Type	Points	Standard
23	MC	1	CCSS.ELA-Literacy.RL.7.4
24	MC	1	CCSS.ELA-Literacy.RL.7.1
25	MC	1	CCSS.ELA-Literacy.RL.7.2
26	MC	1	CCSS.ELA-Literacy.RL.7.6
27	MC	1	CCSS.ELA-Literacy.RL.7.2
28	MC	1	CCSS.ELA-Literacy.RL.7.3
29	MC	1	CCSS.ELA-Literacy.RL.7.4
30	MC	1	CCSS.ELA-Literacy.L.7.5b
31	MC	1	CCSS.ELA-Literacy.RL.7.4
32	MC	1	CCSS.ELA-Literacy.RL.7.2
33	MC	1	CCSS.ELA-Literacy.RL.7.3
34	MC	1	CCSS.ELA-Literacy.RL.7.6
35	MC	1	CCSS.ELA-Literacy.RL.7.3
36	MC	1	CCSS.ELA-Literacy.L.7.4a
37	MC	1	CCSS.ELA-Literacy.RI.7.8
38	MC	1	CCSS.ELA-Literacy.RI.7.1
39	MC	1	CCSS.ELA-Literacy.RI.7.4
40	MC	1	CCSS.ELA-Literacy.RI.7.3
41	MC	1	CCSS.ELA-Literacy.RI.7.2
42	MC	1	CCSS.ELA-Literacy.RI.7.5
43	CR	2	CCSS.ELA-Literacy.RI.7.1
44	CR	2	CCSS.ELA-Literacy.RI.7.8
45	CR	4	CCSS.ELA-Literacy.RL.7.2
46	CR	2	CCSS.ELA-Literacy.RL.7.5
47	CR	2	CCSS.ELA-Literacy.RL.7.3
48	CR	2	CCSS.ELA-Literacy.RI.7.2
49	CR	2	CCSS.ELA-Literacy.RI.7.5
50	CR	2	CCSS.ELA-Literacy.RI.7.1
51	CR	4	CCSS.ELA-Literacy.RI.7.6

Table G6. ELA Grade 8 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.ELA-Literacy.RL.8.4
2	MC	1	CCSS.ELA-Literacy.RL.8.3
3	MC	1	CCSS.ELA-Literacy.RL.8.1
4	MC	1	CCSS.ELA-Literacy.RL.8.3
5	MC	1	CCSS.ELA-Literacy.RL.8.3
6	MC	1	CCSS.ELA-Literacy.RL.8.2

Item	Type	Points	Standard
7	MC	1	CCSS.ELA-Literacy.RL.8.1
15	MC	1	CCSS.ELA-Literacy.RL.8.3
16	MC	1	CCSS.ELA-Literacy.RL.8.5
17	MC	1	CCSS.ELA-Literacy.RL.8.1
18	MC	1	CCSS.ELA-Literacy.RL.8.6
19	MC	1	CCSS.ELA-Literacy.RL.8.3
20	MC	1	CCSS.ELA-Literacy.L.8.4a
21	MC	1	CCSS.ELA-Literacy.RL.8.2
22	MC	1	CCSS.ELA-Literacy.RI.8.4
23	MC	1	CCSS.ELA-Literacy.RI.8.8
24	MC	1	CCSS.ELA-Literacy.RI.8.4
25	MC	1	CCSS.ELA-Literacy.RI.8.5
26	MC	1	CCSS.ELA-Literacy.RI.8.1
27	MC	1	CCSS.ELA-Literacy.RI.8.3
28	MC	1	CCSS.ELA-Literacy.RI.8.3
29	MC	1	CCSS.ELA-Literacy.RI.8.2
30	MC	1	CCSS.ELA-Literacy.RI.8.5
31	MC	1	CCSS.ELA-Literacy.RI.8.1
32	MC	1	CCSS.ELA-Literacy.RI.8.8
33	MC	1	CCSS.ELA-Literacy.RI.8.8
34	MC	1	CCSS.ELA-Literacy.RI.8.5
35	MC	1	CCSS.ELA-Literacy.RI.8.3
36	MC	1	CCSS.ELA-Literacy.RI.8.1
37	MC	1	CCSS.ELA-Literacy.RI.8.3
38	MC	1	CCSS.ELA-Literacy.RI.8.1
39	MC	1	CCSS.ELA-Literacy.RI.8.2
40	MC	1	CCSS.ELA-Literacy.RI.8.5
41	MC	1	CCSS.ELA-Literacy.RI.8.8
42	MC	1	CCSS.ELA-Literacy.RI.8.6
43	CR	2	CCSS.ELA-Literacy.RI.8.6
44	CR	2	CCSS.ELA-Literacy.RI.8.2
45	CR	4	CCSS.ELA-Literacy.RI.8.2
46	CR	2	CCSS.ELA-Literacy.RI.8.5
47	CR	2	CCSS.ELA-Literacy.RI.8.1
48	CR	2	CCSS.ELA-Literacy.RL.8.2
49	CR	2	CCSS.ELA-Literacy.RL.8.4
50	CR	2	CCSS.ELA-Literacy.RL.8.5
51	CR	4	CCSS.ELA-Literacy.RL.8.3

Table G7. Mathematics Grade 3 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.3.NF.A.1
2	MC	1	CCSS.Math.Content.3.OA.A.4
3	MC	1	CCSS.Math.Content.3.MD.C.6
5	MC	1	CCSS.Math.Content.3.MD.B.3
6	MC	1	CCSS.Math.Content.3.OA.D.8
7	MC	1	CCSS.Math.Content.3.NBT.A.1
8	MC	1	CCSS.Math.Content.3.OA.A.2
9	MC	1	CCSS.Math.Content.3.NF.A.2b
10	MC	1	CCSS.Math.Content.3.OA.A.3
12	MC	1	CCSS.Math.Content.3.OA.B.5
13	MC	1	CCSS.Math.Content.3.MD.B.3
15	MC	1	CCSS.Math.Content.3.OA.D.8
16	MC	1	CCSS.Math.Content.3.NF.A.1
17	MC	1	CCSS.Math.Content.3.OA.D.8
18	MC	1	CCSS.Math.Content.3.G.A.2
20	MC	1	CCSS.Math.Content.3.NF.A.2a
21	MC	1	CCSS.Math.Content.3.OA.B.6
22	MC	1	CCSS.Math.Content.3.NF.A.3a
23	MC	1	CCSS.Math.Content.3.MD.C.5b
24	MC	1	CCSS.Math.Content.3.OA.A.4
25	MC	1	CCSS.Math.Content.3.NF.A.1
26	MC	1	CCSS.Math.Content.3.OA.A.2
27	MC	1	CCSS.Math.Content.3.MD.C.7a
29	MC	1	CCSS.Math.Content.3.OA.B.5
30	MC	1	CCSS.Math.Content.3.NBT.A.1
31	MC	1	CCSS.Math.Content.3.OA.D.8
32	MC	1	CCSS.Math.Content.3.OA.A.3
33	MC	1	CCSS.Math.Content.3.OA.D.8
35	MC	1	CCSS.Math.Content.3.MD.A.1
36	MC	1	CCSS.Math.Content.3.OA.B.6
37	MC	1	CCSS.Math.Content.3.NF.A.3b
38	MC	1	CCSS.Math.Content.3.NF.A.3d
39	MC	1	CCSS.Math.Content.3.MD.C.7b
40	MC	1	CCSS.Math.Content.3.OA.A.3
41	MC	1	CCSS.Math.Content.3.NF.A.2b
43	MC	1	CCSS.Math.Content.3.MD.C.6

Item	Type	Points	Standard
44	MC	1	CCSS.Math.Content.3.OA.D.8
45	CR	2	CCSS.Math.Content.3.NF.A.3d
46	CR	2	CCSS.Math.Content.3.MD.A.2
47	CR	2	CCSS.Math.Content.3.G.A.2
48	CR	2	CCSS.Math.Content.3.OA.D.9
49	CR	2	CCSS.Math.Content.3.NBT.A.3
50	CR	3	CCSS.Math.Content.3.OA.A.3
51	CR	3	CCSS.Math.Content.3.MD.C.7d
52	CR	3	CCSS.Math.Content.3.OA.D.8

Table G8. Mathematics Grade 4 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.4.NBT.A.3
2	MC	1	CCSS.Math.Content.4.MD.C.6
3	MC	1	CCSS.Math.Content.4.NF.B.3a
4	MC	1	CCSS.Math.Content.4.G.A.1
5	MC	1	CCSS.Math.Content.4.NBT.B.5
6	MC	1	CCSS.Math.Content.4.NF.A.1
7	MC	1	CCSS.Math.Content.4.MD.C.5a
8	MC	1	CCSS.Math.Content.4.MD.B.4
10	MC	1	CCSS.Math.Content.4.NBT.B.6
11	MC	1	CCSS.Math.Content.4.MD.A.3
12	MC	1	CCSS.Math.Content.4.OA.A.3
14	MC	1	CCSS.Math.Content.4.NBT.B.6
15	MC	1	CCSS.Math.Content.3.G.A.1
16	MC	1	CCSS.Math.Content.4.MD.C.7
17	MC	1	CCSS.Math.Content.4.NF.A.1
18	MC	1	CCSS.Math.Content.4.NF.B.3a
21	MC	1	CCSS.Math.Content.4.MD.C.7
22	MC	1	CCSS.Math.Content.4.NBT.B.6
23	MC	1	CCSS.Math.Content.4.OA.B.4
24	MC	1	CCSS.Math.Content.4.MD.B.4
25	MC	1	CCSS.Math.Content.4.NBT.B.5
26	MC	1	CCSS.Math.Content.4.OA.A.1
27	MC	1	CCSS.Math.Content.4.NF.A.1
28	MC	1	CCSS.Math.Content.3.MD.D.8
29	MC	1	CCSS.Math.Content.4.G.A.3
30	MC	1	CCSS.Math.Content.4.NBT.B.6

Item	Type	Points	Standard
31	MC	1	CCSS.Math.Content.4.OA.A.2
32	MC	1	CCSS.Math.Content.4.NF.B.4c
33	MC	1	CCSS.Math.Content.4.MD.C.6
35	MC	1	CCSS.Math.Content.4.NF.A.2
36	MC	1	CCSS.Math.Content.4.G.A.1
37	MC	1	CCSS.Math.Content.4.MD.C.5b
38	MC	1	CCSS.Math.Content.4.OA.C.5
40	MC	1	CCSS.Math.Content.4.NBT.B.5
41	MC	1	CCSS.Math.Content.4.NF.A.1
43	MC	1	CCSS.Math.Content.4.NF.B.4c
44	MC	1	CCSS.Math.Content.4.NBT.B.6
45	MC	1	CCSS.Math.Content.4.NF.B.3c
46	CR	2	CCSS.Math.Content.4.NF.B.3d
47	CR	2	CCSS.Math.Content.4.NBT.B.5
48	CR	2	CCSS.Math.Content.4.G.A.1
49	CR	2	CCSS.Math.Content.4.NF.A.2
50	CR	2	CCSS.Math.Content.4.MD.A.3
51	CR	2	CCSS.Math.Content.4.NBT.A.2
52	CR	3	CCSS.Math.Content.4.OA.A.3
53	CR	3	CCSS.Math.Content.4.NBT.B.5
54	CR	3	CCSS.Math.Content.4.NF.B.4b
55	CR	3	CCSS.Math.Content.4.OA.A.2

Table G9. Mathematics Grade 5 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.5.OA.A.1
2	MC	1	CCSS.Math.Content.5.NF.A.1
3	MC	1	CCSS.Math.Content.5.MD.C.4
4	MC	1	CCSS.Math.Content.4.NF.C.6
6	MC	1	CCSS.Math.Content.5.NBT.A.3a
7	MC	1	CCSS.Math.Content.5.OA.A.2
8	MC	1	CCSS.Math.Content.5.NBT.A.1
9	MC	1	CCSS.Math.Content.5.MD.C.5b
11	MC	1	CCSS.Math.Content.5.NF.B.5a
12	MC	1	CCSS.Math.Content.5.NBT.B.7
13	MC	1	CCSS.Math.Content.5.NF.B.7a
14	MC	1	CCSS.Math.Content.5.MD.A.1
15	MC	1	CCSS.Math.Content.5.NBT.B.7

Item	Type	Points	Standard
16	MC	1	CCSS.Math.Content.5.NF.B.3
17	MC	1	CCSS.Math.Content.5.MD.C.3a
18	MC	1	CCSS.Math.Content.5.OA.A.2
21	MC	1	CCSS.Math.Content.5.NF.A.2
22	MC	1	CCSS.Math.Content.5.NBT.A.4
23	MC	1	CCSS.Math.Content.5.MD.C.3b
24	MC	1	CCSS.Math.Content.5.NBT.A.2
25	MC	1	CCSS.Math.Content.5.G.B.3
26	MC	1	CCSS.Math.Content.5.NBT.A.3b
27	MC	1	CCSS.Math.Content.4.MD.A.1
28	MC	1	CCSS.Math.Content.5.MD.B.2
29	MC	1	CCSS.Math.Content.5.NF.B.5b
31	MC	1	CCSS.Math.Content.5.MD.C.4
32	MC	1	CCSS.Math.Content.5.NF.B.4
33	MC	1	CCSS.Math.Content.5.MD.C.5a
35	MC	1	CCSS.Math.Content.5.OA.A.1
36	MC	1	CCSS.Math.Content.5.MD.C.3a
37	MC	1	CCSS.Math.Content.5.NF.B.7
38	MC	1	CCSS.Math.Content.5.NBT.A.2
39	MC	1	CCSS.Math.Content.5.G.B.3
41	MC	1	CCSS.Math.Content.5.NF.B.5a
42	MC	1	CCSS.Math.Content.5.MD.A.1
43	MC	1	CCSS.Math.Content.5.NF.B.7a
44	MC	1	CCSS.Math.Content.5.MD.C.3b
45	MC	1	CCSS.Math.Content.5.NBT.B.7
46	CR	2	CCSS.Math.Content.5.MD.A.1
47	CR	2	CCSS.Math.Content.5.NF.A.2
48	CR	2	CCSS.Math.Content.5.NBT.B.6
49	CR	2	CCSS.Math.Content.5.NF.A.2
50	CR	2	CCSS.Math.Content.5.NBT.B.7
51	CR	2	CCSS.Math.Content.5.MD.C.5b
52	CR	3	CCSS.Math.Content.5.NF.B.6
53	CR	3	CCSS.Math.Content.5.NBT.B.6
54	CR	3	CCSS.Math.Content.5.NF.B.6
55	CR	3	CCSS.Math.Content.5.NF.A.2

Table G10. Mathematics Grade 6 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.6.RP.A.3a
2	MC	1	CCSS.Math.Content.6.EE.B.5
3	MC	1	CCSS.Math.Content.6.NS.A.1
4	MC	1	CCSS.Math.Content.6.EE.A.1
6	MC	1	CCSS.Math.Content.6.NS.B.4
7	MC	1	CCSS.Math.Content.6.G.A.1
9	MC	1	CCSS.Math.Content.6.G.A.1
10	MC	1	CCSS.Math.Content.5.G.A.2
11	MC	1	CCSS.Math.Content.6.RP.A.2
13	MC	1	CCSS.Math.Content.6.EE.C.9
14	MC	1	CCSS.Math.Content.6.RP.A.3a
15	MC	1	CCSS.Math.Content.6.EE.A.2a
16	MC	1	CCSS.Math.Content.6.EE.B.8
17	MC	1	CCSS.Math.Content.6.G.A.4
18	MC	1	CCSS.Math.Content.6.RP.A.3c
19	MC	1	CCSS.Math.Content.6.EE.B.6
20	MC	1	CCSS.Math.Content.6.NS.A.1
21	MC	1	CCSS.Math.Content.5.G.A.2
23	MC	1	CCSS.Math.Content.6.EE.A.3
24	MC	1	CCSS.Math.Content.6.RP.A.3c
25	MC	1	CCSS.Math.Content.6.EE.A.4
26	MC	1	CCSS.Math.Content.6.NS.B.4
27	MC	1	CCSS.Math.Content.6.RP.A.3c
28	MC	1	CCSS.Math.Content.6.NS.B.4
29	MC	1	CCSS.Math.Content.6.EE.A.2c
30	MC	1	CCSS.Math.Content.6.NS.C.6a
32	MC	1	CCSS.Math.Content.6.EE.A.3
33	MC	1	CCSS.Math.Content.6.EE.B.6
34	MC	1	CCSS.Math.Content.6.EE.C.9
35	MC	1	CCSS.Math.Content.6.RP.A.3
36	MC	1	CCSS.Math.Content.6.G.A.3
37	MC	1	CCSS.Math.Content.6.EE.B.6
38	MC	1	CCSS.Math.Content.6.NS.C.7a
39	MC	1	CCSS.Math.Content.6.EE.B.5
40	MC	1	CCSS.Math.Content.6.G.A.2
41	MC	1	CCSS.Math.Content.6.RP.A.2

Item	Type	Points	Standard
42	MC	1	CCSS.Math.Content.6.G.A.4
43	MC	1	CCSS.Math.Content.6.EE.A.4
44	MC	1	CCSS.Math.Content.6.NS.B.4
45	MC	1	CCSS.Math.Content.6.EE.A.3
46	MC	1	CCSS.Math.Content.5.G.A.1
47	MC	1	CCSS.Math.Content.6.NS.A.1
50	MC	1	CCSS.Math.Content.6.G.A.1
51	MC	1	CCSS.Math.Content.6.EE.B.7
52	CR	2	CCSS.Math.Content.6.RP.A.3b
53	CR	2	CCSS.Math.Content.6.EE.A.2c
54	CR	2	CCSS.Math.Content.6.RP.A.3d
55	CR	2	CCSS.Math.Content.6.EE.B.7
56	CR	2	CCSS.Math.Content.6.RP.A.3c
57	CR	2	CCSS.Math.Content.6.EE.A.3
58	CR	3	CCSS.Math.Content.6.EE.B.7
59	CR	3	CCSS.Math.Content.6.RP.A.3b
60	CR	3	CCSS.Math.Content.6.G.A.2
61	CR	3	CCSS.Math.Content.6.NS.C.5

Table G11. Mathematics Grade 7 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.7.NS.A.1b
2	MC	1	CCSS.Math.Content.7.EE.B.4a
3	MC	1	CCSS.Math.Content.7.RP.A.2a
4	MC	1	CCSS.Math.Content.7.RP.A.3
6	MC	1	CCSS.Math.Content.7.NS.A.3
8	MC	1	CCSS.Math.Content.7.G.B.4
9	MC	1	CCSS.Math.Content.7.EE.A.2
10	MC	1	CCSS.Math.Content.7.NS.A.1c
11	MC	1	CCSS.Math.Content.7.EE.B.4b
12	MC	1	CCSS.Math.Content.7.RP.A.1
13	MC	1	CCSS.Math.Content.7.EE.A.1
14	MC	1	CCSS.Math.Content.7.NS.A.2d
15	MC	1	CCSS.Math.Content.7.EE.B.4a
16	MC	1	CCSS.Math.Content.7.NS.A.2a
18	MC	1	CCSS.Math.Content.7.EE.B.3
19	MC	1	CCSS.Math.Content.7.G.A.1
20	MC	1	CCSS.Math.Content.7.NS.A.2d

Item	Type	Points	Standard
21	MC	1	CCSS.Math.Content.7.EE.A.1
22	MC	1	CCSS.Math.Content.7.SP.C.5
23	MC	1	CCSS.Math.Content.7.NS.A.3
25	MC	1	CCSS.Math.Content.7.NS.A.1a
26	MC	1	CCSS.Math.Content.7.SP.B.4
27	MC	1	CCSS.Math.Content.7.G.A.1
28	MC	1	CCSS.Math.Content.7.SP.C.7a
29	MC	1	CCSS.Math.Content.6.SP.A.3
30	MC	1	CCSS.Math.Content.7.EE.B.4a
32	MC	1	CCSS.Math.Content.7.G.A.1
33	MC	1	CCSS.Math.Content.7.RP.A.1
34	MC	1	CCSS.Math.Content.7.SP.B.3
35	MC	1	CCSS.Math.Content.7.EE.A.1
36	MC	1	CCSS.Math.Content.7.RP.A.3
37	MC	1	CCSS.Math.Content.7.SP.A.1
38	MC	1	CCSS.Math.Content.7.EE.A.2
39	MC	1	CCSS.Math.Content.7.G.B.4
40	MC	1	CCSS.Math.Content.7.EE.A.1
41	MC	1	CCSS.Math.Content.7.RP.A.3
42	MC	1	CCSS.Math.Content.7.EE.B.4b
43	MC	1	CCSS.Math.Content.7.RP.A.2d
44	MC	1	CCSS.Math.Content.7.EE.A.2
45	MC	1	CCSS.Math.Content.7.G.B.4
46	MC	1	CCSS.Math.Content.7.NS.A.3
47	MC	1	CCSS.Math.Content.7.RP.A.3
50	MC	1	CCSS.Math.Content.7.NS.A.2c
51	MC	1	CCSS.Math.Content.7.SP.C.6
52	CR	2	CCSS.Math.Content.7.EE.B.3
53	CR	2	CCSS.Math.Content.7.EE.B.3
54	CR	2	CCSS.Math.Content.7.RP.A.3
55	CR	2	CCSS.Math.Content.7.SP.C.6
56	CR	2	CCSS.Math.Content.7.NS.A.3
57	CR	2	CCSS.Math.Content.7.RP.A.2b
58	CR	3	CCSS.Math.Content.7.NS.A.3
59	CR	3	CCSS.Math.Content.7.RP.A.2
60	CR	3	CCSS.Math.Content.7.EE.B.3
61	CR	3	CCSS.Math.Content.7.RP.A.3

Table G12. Mathematics Grade 8 Operational Item Map

Item	Type	Points	Standard
1	MC	1	CCSS.Math.Content.8.EE.A.3
2	MC	1	CCSS.Math.Content.8.F.B.5
3	MC	1	CCSS.Math.Content.7.G.A.3
4	MC	1	CCSS.Math.Content.8.F.B.4
5	MC	1	CCSS.Math.Content.8.G.A.2
6	MC	1	CCSS.Math.Content.8.F.A.3
7	MC	1	CCSS.Math.Content.8.EE.C.8a
8	MC	1	CCSS.Math.Content.8.F.A.2
9	MC	1	CCSS.Math.Content.8.EE.C.7b
10	MC	1	CCSS.Math.Content.8.SP.A.3
12	MC	1	CCSS.Math.Content.8.EE.A.3
14	MC	1	CCSS.Math.Content.8.G.A.5
15	MC	1	CCSS.Math.Content.8.EE.B.5
17	MC	1	CCSS.Math.Content.8.F.B.4
18	MC	1	CCSS.Math.Content.8.EE.C.8c
19	MC	1	CCSS.Math.Content.8.SP.A.4
20	MC	1	CCSS.Math.Content.8.G.A.4
22	MC	1	CCSS.Math.Content.8.EE.C.7a
23	MC	1	CCSS.Math.Content.8.SP.A.2
24	MC	1	CCSS.Math.Content.8.EE.A.1
25	MC	1	CCSS.Math.Content.8.F.A.1
26	MC	1	CCSS.Math.Content.8.G.A.3
27	MC	1	CCSS.Math.Content.8.F.B.5
28	MC	1	CCSS.Math.Content.8.SP.A.3
29	MC	1	CCSS.Math.Content.8.G.A.2
30	MC	1	CCSS.Math.Content.7.G.B.6
31	MC	1	CCSS.Math.Content.8.EE.A.1
32	MC	1	CCSS.Math.Content.8.G.C.9
34	MC	1	CCSS.Math.Content.8.G.A.1
35	MC	1	CCSS.Math.Content.8.EE.C.8b
36	MC	1	CCSS.Math.Content.8.SP.A.1
37	MC	1	CCSS.Math.Content.8.EE.C.7b
38	MC	1	CCSS.Math.Content.8.F.A.1
39	MC	1	CCSS.Math.Content.8.EE.C.8c
40	MC	1	CCSS.Math.Content.8.F.B.4
42	MC	1	CCSS.Math.Content.8.EE.A.4
43	MC	1	CCSS.Math.Content.8.F.A.2

Item	Type	Points	Standard
44	MC	1	CCSS.Math.Content.8.G.A.4
45	MC	1	CCSS.Math.Content.8.EE.A.4
46	MC	1	CCSS.Math.Content.8.F.A.3
47	MC	1	CCSS.Math.Content.8.EE.C.7a
48	MC	1	CCSS.Math.Content.8.EE.C.8b
50	MC	1	CCSS.Math.Content.8.EE.B.5
51	MC	1	CCSS.Math.Content.8.G.A.4
52	CR	2	CCSS.Math.Content.8.EE.C.7b
53	CR	2	CCSS.Math.Content.8.G.C.9
54	CR	2	CCSS.Math.Content.8.EE.C.8b
55	CR	2	CCSS.Math.Content.8.EE.B.6
56	CR	2	CCSS.Math.Content.8.F.A.2
57	CR	2	CCSS.Math.Content.8.F.B.4
58	CR	3	CCSS.Math.Content.8.G.C.9
59	CR	3	CCSS.Math.Content.8.EE.B.5
60	CR	3	CCSS.Math.Content.8.EE.C.8c
61	CR	3	CCSS.Math.Content.8.F.A.3

Appendix H: ELA Short-Response Rubric

2-Point Rubric–Short Response

Score	Response Features
2 Point	<p>The features of a 2-point response are:</p> <ul style="list-style-type: none"> • Valid inferences and/or claims from the text where required by the prompt • Evidence of analysis of the text where required by the prompt • Relevant facts, definitions, concrete details, and/or other information from the text to develop response according to the requirements of the prompt • Sufficient number of facts, definitions, concrete details, and/or other information from the text as required by the prompt • Complete sentences where errors do not affect readability
1 Point	<p>The features of a 1-point response are:</p> <ul style="list-style-type: none"> • A mostly literal recounting of events or details from the text as required by the prompt • Some relevant facts, definitions, concrete details, and/or other information from the text to develop response according to the requirements of the prompt • Incomplete sentences or bullets
0 Point*	<p>The features of a 0-point response are:</p> <ul style="list-style-type: none"> • A response that does not address any of the requirements of the prompt or is totally inaccurate • A response that is not written in English • A response that is unintelligible or indecipherable

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

- If the prompt requires two texts and the student only references one text, the response can be scored no higher than a 1.

Appendix I: ELA Extended-Response Rubric

New York State Grade 3 Expository Writing Evaluation Rubric

CRITERIA	CCLS	SCORE				
		4 Essays at this level:	3 Essays at this level:	2 Essays at this level:	1 Essays at this level:	0* Essays at this level:
CONTENT AND ANALYSIS: the extent to which the essay conveys ideas and information clearly and accurately in order to support analysis of topics or text	W.2, R.1–9	–clearly introduce a topic in a manner that follows logically from the task and purpose –demonstrate comprehension and analysis of the text	–clearly introduce a topic in a manner that follows from the task and purpose –demonstrate grade-appropriate comprehension of the text	–introduce a topic in a manner that follows generally from the task and purpose –demonstrate a confused comprehension of the text	–introduce a topic in a manner that does not logically follow from the task and purpose –demonstrate little understanding of the text	–demonstrate a lack of comprehension of the text or task
COMMAND OF EVIDENCE: the extent to which the essay presents evidence from the provided text to support analysis and reflection	W.2 R.1–8	–develop the topic with relevant, well-chosen facts, definitions, and details throughout the essay	–develop the topic with relevant facts, definitions, and details throughout the essay	–partially develop the topic of the essay with the use of some textual evidence, some of which may be irrelevant	–demonstrate an attempt to use evidence, but only develop ideas with minimal, occasional evidence which is generally invalid or irrelevant	–provide no evidence or provide evidence that is completely irrelevant
COHERENCE, ORGANIZATION, AND STYLE: the extent to which the essay logically organizes complex ideas, concepts, and information using formal style and precise language	W.2 L.3 L.6	–clearly and consistently group related information together –skillfully connect ideas within categories of information using linking words and phrases –provide a concluding statement that follows clearly from the topic and information presented	–generally group related information together –connect ideas within categories of information using linking words and phrases –provide a concluding statement that follows from the topic and information presented	–exhibit some attempt to group related information together –inconsistently connect ideas using some linking words and phrases –provide a concluding statement that follows generally from the topic and information presented	–exhibit little attempt at organization –lack the use of linking words and phrases –provide a concluding statement that is illogical or unrelated to the topic and information presented	–exhibit no evidence of organization –do not provide a concluding statement
CONTROL OF CONVENTIONS: the extent to which the essay demonstrates command of the conventions of standard English grammar, usage, capitalization, punctuation, and spelling	W.2 L.1 L.2	–demonstrate grade-appropriate command of conventions, with few errors	–demonstrate grade-appropriate command of conventions, with occasional errors that do not hinder comprehension	–demonstrate emerging command of conventions, with some errors that may hinder comprehension	–demonstrate a lack of command of conventions, with frequent errors that hinder comprehension	–are minimal, making assessment of conventions unreliable

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

- If the student writes only a personal response and makes no reference to the text(s), the response can be scored no higher than a 1.
- Responses totally unrelated to the topic, illegible, or incoherent should be given a 0.
- A response totally copied from the text(s) with no original student writing should be scored a 0.

New York State Grade 4-5 Expository Writing Evaluation Rubric

CRITERIA	CCLS	SCORE				
		4 Essays at this level:	3 Essays at this level:	2 Essays at this level:	1 Essays at this level:	0* Essays at this level:
CONTENT AND ANALYSIS: the extent to which the essay conveys ideas and information clearly and accurately in order to support an analysis of topics or texts	W.2 R.1–9	– clearly introduce a topic in a manner that follows logically from the task and purpose – demonstrate insightful comprehension and analysis of the text(s)	– clearly introduce a topic in a manner that follows from the task and purpose – demonstrate grade-appropriate comprehension and analysis of the text(s)	– introduce a topic in a manner that follows generally from the task and purpose – demonstrate a literal comprehension of the text(s)	– introduce a topic in a manner that does not logically follow from the task and purpose – demonstrate little understanding of the text(s)	– demonstrate a lack of comprehension of the text(s) or task
COMMAND OF EVIDENCE: the extent to which the essay presents evidence from the provided texts to support analysis and reflection	W.2 W.9 R.1–9	– develop the topic with relevant, well-chosen facts, definitions, concrete details, quotations, or other information and examples from the text(s) – sustain the use of varied, relevant evidence	– develop the topic with relevant facts, definitions, details, quotations, or other information and examples from the text(s) – sustain the use of relevant evidence, with some lack of variety	– partially develop the topic of the essay with the use of some textual evidence, some of which may be irrelevant – use relevant evidence with inconsistency	– demonstrate an attempt to use evidence, but only develop ideas with minimal, occasional evidence which is generally invalid or irrelevant	– provide no evidence or provide evidence that is completely irrelevant
COHERENCE, ORGANIZATION, AND STYLE: the extent to which the essay logically organizes complex ideas, concepts, and information using formal style and precise language	W.2 L.3 L.6	– exhibit clear, purposeful organization – skillfully link ideas using grade-appropriate words and phrases – use grade-appropriate, stylistically sophisticated language and domain-specific vocabulary – provide a concluding statement that follows clearly from the topic and information presented	– exhibit clear organization – link ideas using grade-appropriate words and phrases – use grade-appropriate precise language and domain-specific vocabulary – provide a concluding statement that follows from the topic and information presented	– exhibit some attempt at organization – inconsistently link ideas using words and phrases – inconsistently use appropriate language and domain-specific vocabulary – provide a concluding statement that follows generally from the topic and information presented	– exhibit little attempt at organization, or attempts to organize are irrelevant to the task – lack the use of linking words and phrases – use language that is imprecise or inappropriate for the text(s) and task – provide a concluding statement that is illogical or unrelated to the topic and information presented	– exhibit no evidence of organization – exhibit no use of linking words and phrases – use language that is predominantly incoherent or copied directly from the text(s) – do not provide a concluding statement
CONTROL OF CONVENTIONS: the extent to which the essay demonstrates command of the conventions of standard English grammar, usage, capitalization, punctuation, and spelling	W.2 L.1 L.2	– demonstrate grade-appropriate command of conventions, with few errors	– demonstrate grade-appropriate command of conventions, with occasional errors that do not hinder comprehension	– demonstrate emerging command of conventions, with some errors that may hinder comprehension	– demonstrate a lack of command of conventions, with frequent errors that hinder comprehension	– are minimal, making assessment of conventions unreliable

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

- If the prompt requires two texts and the student only references one text, the response can be scored no higher than a 2.
- If the student writes only a personal response and makes no reference to the text(s), the response can be scored no higher than a 1.
- Responses totally unrelated to the topic, illegible, or incoherent should be given a 0.
- A response totally copied from the text(s) with no original student writing should be scored a 0.

New York State Grade 6-8 Expository Writing Evaluation Rubric

CRITERIA	CCLS	SCORE				
		4 Essays at this level:	3 Essays at this level:	2 Essays at this level:	1 Essays at this level:	0* Essays at this level:
CONTENT AND ANALYSIS: the extent to which the essay conveys complex ideas and information clearly and accurately in order to support claims in an analysis of topics or texts	W.2, R.1-9	–clearly introduce a topic in a manner that is compelling and follows logically from the task and purpose –demonstrate insightful analysis of the text(s)	– clearly introduce a topic in a manner that follows from the task and purpose –demonstrate grade-appropriate analysis of the text(s)	–introduce a topic in a manner that follows generally from the task and purpose –demonstrate a literal comprehension of the text(s)	–introduce a topic in a manner that does not logically follow from the task and purpose –demonstrate little understanding of the text(s)	–demonstrate a lack of comprehension of the text(s) or task
COMMAND OF EVIDENCE: the extent to which the essay presents evidence from the provided texts to support analysis and reflection	W.9, R.1-9	–develop the topic with relevant, well-chosen facts, definitions, concrete details, quotations, or other information and examples from the text(s) –sustain the use of varied, relevant evidence	–develop the topic with relevant facts, definitions, details, quotations, or other information and examples from the text(s) –sustain the use of relevant evidence, with some lack of variety	–partially develop the topic of the essay with the use of some textual evidence, some of which may be irrelevant –use relevant evidence with inconsistency	–demonstrate an attempt to use evidence, but only develop ideas with minimal, occasional evidence which is generally invalid or irrelevant	–provide no evidence or provide evidence that is completely irrelevant
COHERENCE, ORGANIZATION, AND STYLE: the extent to which the essay logically organizes complex ideas, concepts, and information using formal style and precise language	W.2, L.3, L.6	–exhibit clear organization, with the skillful use of appropriate and varied transitions to create a unified whole and enhance meaning –establish and maintain a formal style, using grade-appropriate, stylistically sophisticated language and domain-specific vocabulary with a notable sense of voice –provide a concluding statement or section that is compelling and follows clearly from the topic and information presented	–exhibit clear organization, with the use of appropriate transitions to create a unified whole –establish and maintain a formal style using precise language and domain-specific vocabulary –provide a concluding statement or section that follows from the topic and information presented	–exhibit some attempt at organization, with inconsistent use of transitions –establish but fail to maintain a formal style, with inconsistent use of language and domain-specific vocabulary –provide a concluding statement or section that follows generally from the topic and information presented	–exhibit little attempt at organization, or attempts to organize are irrelevant to the task –lack a formal style, using language that is imprecise or inappropriate for the text(s) and task –provide a concluding statement or section that is illogical or unrelated to the topic and information presented	–exhibit no evidence of organization –use language that is predominantly incoherent or copied directly from the text(s) –do not provide a concluding statement or section
CONTROL OF CONVENTIONS: the extent to which the essay demonstrates command of the conventions of standard English grammar, usage, capitalization, punctuation, and spelling	W.2, L.1, L.2	–demonstrate grade-appropriate command of conventions, with few errors	–demonstrate grade-appropriate command of conventions, with occasional errors that do not hinder comprehension	–demonstrate emerging command of conventions, with some errors that may hinder comprehension	–demonstrate a lack of command of conventions, with frequent errors that hinder comprehension	–are minimal, making assessment of conventions unreliable

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

- If the prompt requires two texts and the student only references one text, the response can be scored no higher than a 2.
- If the student writes only a personal response and makes no reference to the text(s), the response can be scored no higher than a 1.
- Responses totally unrelated to the topic, illegible, or incoherent should be given a 0.
- A response totally copied from the text(s) with no original student writing should be scored a 0.

Appendix J: Mathematics Short-Response Rubric

2-Point Holistic Rubric

2 Points	<p>A two-point response includes the correct solution to the question and demonstrates a thorough understanding of the mathematical concepts and/or procedures in the task.</p> <p>This response:</p> <ul style="list-style-type: none"> • indicates that the student has completed the task correctly, using mathematically sound procedures • contains sufficient work to demonstrate a thorough understanding of the mathematical concepts and/or procedures • may contain inconsequential errors that do not detract from the correct solution and the demonstration of a thorough understanding
1 Point	<p>A one-point response demonstrates only a partial understanding of the mathematical concepts and/or procedures in the task.</p> <p>This response:</p> <ul style="list-style-type: none"> • correctly addresses only some elements of the task • may contain an incorrect solution but applies a mathematically appropriate process • may contain the correct solution but required work is incomplete
0 Points*	<p>A zero-point response is incorrect, irrelevant, incoherent, or contains a correct solution obtained using an obviously incorrect procedure. Although some elements may contain correct mathematical procedures, holistically they are not sufficient to demonstrate even a limited understanding of the mathematical concepts embodied in the task.</p>

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

Appendix K: Mathematics Extended-Response Rubric

3-Point Holistic Rubric

3 Points	<p>A three-point response includes the correct solution(s) to the question and demonstrates a thorough understanding of the mathematical concepts and/or procedures in the task.</p> <p>This response:</p> <ul style="list-style-type: none"> • indicates that the student has completed the task correctly, using mathematically sound procedures • contains sufficient work to demonstrate a thorough understanding of the mathematical concepts and/or procedures • may contain inconsequential errors that do not detract from the correct solution(s) and the demonstration of a thorough understanding
2 Points	<p>A two-point response demonstrates a partial understanding of the mathematical concepts and/or procedures in the task.</p> <p>This response:</p> <ul style="list-style-type: none"> • appropriately addresses most, but not all, aspects of the task using mathematically sound procedures • may contain an incorrect solution but provides sound procedures, reasoning, and/or explanations • may reflect some minor misunderstanding of the underlying mathematical concepts and/or procedures
1 Point	<p>A one-point response demonstrates only a limited understanding of the mathematical concepts and/or procedures in the task.</p> <p>This response:</p> <ul style="list-style-type: none"> • may address some elements of the task correctly but reaches an inadequate solution and/or provides reasoning that is faulty or incomplete • exhibits multiple flaws related to misunderstanding of important aspects of the task, misuse of mathematical procedures, or faulty mathematical reasoning • reflects a lack of essential understanding of the underlying mathematical concepts • may contain the correct solution(s) but required work is limited
0 Points*	<p>A zero-point response is incorrect, irrelevant, incoherent, or contains a correct solution obtained using an obviously incorrect procedure. Although some elements may contain correct mathematical procedures, holistically they are not sufficient to demonstrate even a limited understanding of the mathematical concepts embodied in the task.</p>

* Condition Code A is applied whenever a student who is present for a test session leaves an entire constructed-response question in that session completely blank (no response attempted).

Appendix L: Factor Analysis Results for Select Subgroups

As described in Section 3: Validity, a principal components factor analysis was conducted on the Grades 3–8 ELA and Mathematics Tests data. The analyses were conducted for the total population of students and select subgroups: ELL, SWD, SUA, SWD/SUA students using disability accommodations, and ELL students using ELL-related accommodations (ELL & SUA). Tables L1 and L2 contain the results of factor analysis on the subpopulation data for the Grades 3–8 ELA and Mathematics Tests, respectively.

Table L1. ELA Grade 3 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL	ELL=Y	6.46	18.99	18.99
		1.49	4.39	23.38
		1.12	3.30	26.68
		1.04	3.06	29.74
		1.03	3.03	32.78
		1.01	2.98	35.75
SWD	All Codes	7.32	21.53	21.53
		1.55	4.57	26.10
		1.08	3.18	29.28
		1.03	3.02	32.30
SUA	All Codes	6.99	20.55	20.55
		1.58	4.65	25.20
		1.09	3.19	28.40
		1.03	3.04	31.44
SWD/SUA	SUA=504 plan codes	6.73	19.80	19.80
		1.61	4.73	24.53
		1.09	3.22	27.75
		1.04	3.07	30.82
		1.00	2.95	33.77
ELL/SUA	SUA & ELL Codes	5.74	16.87	16.87
		1.54	4.52	21.39
		1.20	3.54	24.94
		1.11	3.26	28.19
		1.06	3.11	31.31
		1.04	3.07	34.37
		1.03	3.03	37.41
		1.02	2.99	40.40
		1.01	2.96	43.36

Appendix L: Factor Analysis Results for Select Subgroups

Table L2. ELA Grade 4 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL	ELL=Y	5.30	15.58	15.58
		1.47	4.32	19.90
		1.10	3.23	23.14
		1.08	3.19	26.33
		1.07	3.14	29.47
		1.05	3.08	32.55
		1.03	3.04	35.59
		1.02	2.99	38.58
		1.01	2.97	41.55
SWD	All Codes	6.25	18.37	18.37
		1.50	4.41	22.78
		1.11	3.27	26.06
		1.06	3.12	29.18
		1.02	2.99	32.16
		1.01	2.96	35.12
SUA	All Codes	6.12	17.99	17.99
		1.51	4.43	22.42
		1.12	3.28	25.71
		1.07	3.14	28.85
		1.02	3.00	31.85
		1.01	2.97	34.82
SWD/SUA	SUA=504 plan codes	5.89	17.32	17.32
		1.52	4.46	21.78
		1.13	3.32	25.10
		1.07	3.15	28.25
		1.02	3.01	31.25
		1.02	3.00	34.25
		1.01	2.96	37.21
ELL/SUA	SUA & ELL Codes	4.71	13.87	13.87
		1.50	4.40	18.27
		1.18	3.47	21.74
		1.16	3.43	25.16
		1.12	3.29	28.46
		1.11	3.27	31.73

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/SUA	SUA & ELL Codes	1.09	3.21	34.94
		1.07	3.14	38.08
		1.04	3.07	41.14
		1.04	3.05	44.19

Table L3. ELA Grade 5 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL	ELL=Y	5.45	12.40	12.40
		1.51	3.42	15.82
		1.15	2.61	18.43
		1.14	2.58	21.01
		1.11	2.53	23.54
		1.10	2.50	26.04
		1.08	2.47	28.50
		1.07	2.44	30.95
		1.05	2.39	33.34
		1.05	2.39	35.73
		1.03	2.34	38.08
		1.02	2.31	40.39
		1.00	2.28	42.66
SWD	All Codes	6.59	14.98	14.98
		1.57	3.56	18.54
		1.15	2.60	21.15
		1.10	2.50	23.65
		1.08	2.45	26.10
		1.04	2.37	28.47
		1.03	2.35	30.81
		1.02	2.31	33.12
		1.01	2.30	35.42
SUA	All Codes	6.70	15.22	15.22
		1.58	3.59	18.81
		1.15	2.61	21.42
		1.10	2.51	23.93
		1.08	2.45	26.38

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
SUA	All Codes	1.05	2.38	28.75
		1.03	2.33	31.08
		1.01	2.30	33.38
		1.00	2.28	35.66
SWD/SUA	SUA=504 plan codes	6.31	14.34	14.34
		1.59	3.62	17.95
		1.15	2.61	20.57
		1.10	2.51	23.07
		1.09	2.47	25.54
		1.06	2.40	27.95
		1.04	2.36	30.31
		1.03	2.33	32.64
		1.02	2.31	34.95
		1.01	2.28	37.23
ELL/SUA	SUA & ELL Codes	4.89	11.11	11.11
		1.64	3.72	14.83
		1.24	2.82	17.65
		1.18	2.69	20.34
		1.15	2.62	22.96
		1.14	2.59	25.55
		1.13	2.56	28.11
		1.12	2.55	30.66
		1.08	2.46	33.12
		1.07	2.44	35.56
		1.06	2.41	37.96
		1.05	2.38	40.34
		1.04	2.36	42.70
		1.03	2.35	45.05
		1.02	2.31	47.36
		1.01	2.29	49.64

Table L4. ELA Grade 6 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL	ELL=Y	6.53	14.83	14.83
		1.83	4.16	18.99
		1.18	2.68	21.67
		1.12	2.55	24.22
		1.09	2.48	26.71
		1.07	2.44	29.14
		1.06	2.40	31.54
		1.04	2.37	33.91
		1.04	2.35	36.26
		1.01	2.30	38.56
		1.01	2.29	40.85
SWD	All Codes	7.55	17.17	17.17
		2.13	4.84	22.01
		1.15	2.62	24.63
		1.06	2.40	27.03
		1.05	2.40	29.43
		1.04	2.36	31.79
		1.01	2.30	34.09
SUA	All Codes	7.77	17.67	17.67
		2.16	4.91	22.58
		1.16	2.63	25.20
		1.06	2.42	27.62
		1.05	2.39	30.01
		1.03	2.34	32.36
		1.00	2.28	34.63
SWD/SUA	SUA=504 plan codes	7.34	16.68	16.68
		2.13	4.85	21.53
		1.16	2.63	24.17
		1.07	2.43	26.60
		1.06	2.41	29.01
		1.05	2.38	31.39
		1.01	2.30	33.69
ELL/SUA	SUA & ELL Codes	5.51	12.53	12.53
		1.69	3.83	16.36

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/SUA	SUA & ELL Codes	1.23	2.79	19.14
		1.21	2.75	21.89
		1.18	2.67	24.56
		1.14	2.60	27.16
		1.12	2.55	29.71
		1.11	2.52	32.24
		1.10	2.50	34.73
		1.08	2.45	37.19
		1.05	2.38	39.57
		1.03	2.35	41.92
		1.03	2.34	44.26
		1.03	2.33	46.59
		1.01	2.29	48.88

Table L5. ELA Grade 7 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL	ELL=Y	6.37	14.49	14.49
		1.68	3.81	18.30
		1.24	2.81	21.11
		1.17	2.65	23.76
		1.14	2.59	26.35
		1.11	2.52	28.87
		1.06	2.41	31.28
		1.05	2.39	33.67
		1.05	2.38	36.05
		1.03	2.34	38.39
		1.02	2.33	40.72
		1.01	2.29	43.01
SWD	All Codes	7.01	15.93	15.93
		2.03	4.62	20.55
		1.22	2.78	23.33
		1.10	2.50	25.83
		1.05	2.39	28.21
		1.02	2.32	30.53

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
SWD	All Codes	1.01	2.29	32.82
		1.00	2.28	35.10
SUA	All Codes	7.29	16.57	16.57
		2.07	4.69	21.27
		1.23	2.79	24.06
		1.09	2.49	26.55
		1.05	2.39	28.93
		1.01	2.29	31.22
		1.00	2.28	33.50
SWD/SUA	SUA=504 plan codes	6.78	15.41	15.41
		2.01	4.58	19.99
		1.22	2.76	22.75
		1.11	2.52	25.27
		1.07	2.42	27.69
		1.03	2.34	30.04
		1.02	2.32	32.35
		1.02	2.31	34.67
		1.00	2.27	36.94
ELL/SUA	SUA & ELL Codes	5.21	11.85	11.85
		1.58	3.58	15.43
		1.26	2.86	18.29
		1.22	2.78	21.06
		1.21	2.75	23.81
		1.17	2.65	26.46
		1.14	2.59	29.05
		1.10	2.50	31.55
		1.08	2.46	34.01
		1.08	2.45	36.46
		1.07	2.43	38.89
		1.05	2.39	41.27
		1.03	2.33	43.60
		1.02	2.32	45.93
		1.01	2.30	48.23

Table L6. ELA Grade 8 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL	ELL=Y	6.59	14.97	14.97
		1.70	3.86	18.83
		1.18	2.69	21.52
		1.14	2.59	24.11
		1.13	2.58	26.69
		1.09	2.48	29.18
		1.08	2.45	31.63
		1.06	2.40	34.03
		1.04	2.37	36.40
		1.03	2.35	38.75
		1.02	2.32	41.07
		1.01	2.29	43.36
SWD	All Codes	7.73	17.58	17.58
		2.00	4.55	22.13
		1.22	2.78	24.91
		1.09	2.48	27.39
		1.06	2.41	29.80
		1.03	2.35	32.15
		1.02	2.31	34.46
		1.01	2.29	36.76
SUA	All Codes	8.10	18.41	18.41
		2.00	4.54	22.96
		1.24	2.81	25.77
		1.08	2.46	28.23
		1.04	2.36	30.59
		1.02	2.33	32.92
		1.01	2.29	35.21
		1.00	2.28	37.48
SWD/SUA	SUA=504 plan codes	7.63	17.35	17.35
		1.99	4.52	21.87
		1.22	2.78	24.65
		1.09	2.48	27.14
		1.06	2.42	29.55
		1.04	2.37	31.92

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
SWD/SUA	SWD & SUA codes	1.02	2.31	34.23
		1.01	2.29	36.53
ELL/SUA	SUA & ELL Codes	6.04	13.72	13.72
		1.74	3.95	17.67
		1.30	2.95	20.62
		1.23	2.79	23.41
		1.20	2.72	26.13
		1.14	2.60	28.73
		1.14	2.59	31.32
		1.13	2.57	33.89
		1.09	2.48	36.38
		1.08	2.46	38.84
		1.07	2.43	41.26
		1.05	2.39	43.65
		1.03	2.35	46.00
		1.03	2.33	48.33
		1.01	2.29	50.62

Table L7. Mathematics Grade 3 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL	ELL=Y	10.41	23.13	23.13
		2.07	4.60	27.74
		1.13	2.51	30.24
		1.11	2.48	32.72
		1.05	2.34	35.06
SWD	All Codes	11.01	24.46	24.46
		1.85	4.11	28.57
		1.10	2.45	31.02
		1.08	2.41	33.43
SUA	All Codes	1.05	2.34	35.77
		10.55	23.45	23.45
		1.85	4.10	27.55
		1.12	2.49	30.04
		1.09	2.43	32.47

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
SUA	All Codes	1.06	2.35	34.81
SWD/SUA	SUA=504 plan codes	10.29	22.86	22.86
		1.85	4.11	26.98
		1.13	2.50	29.48
		1.10	2.44	31.92
		1.07	2.37	34.29
ELL/SUA	SUA & ELL Codes	9.38	20.85	20.85
		1.93	4.28	25.13
		1.16	2.59	27.72
		1.13	2.50	30.22
		1.10	2.45	32.67
		1.03	2.29	34.96
		1.01	2.24	37.20
		1.00	2.22	39.42

Table L8. Mathematics Grade 4 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL	ELL=Y	11.11	23.15	23.15
		1.82	3.79	26.93
		1.20	2.50	29.44
		1.10	2.30	31.73
		1.05	2.20	33.93
SWD	All Codes	11.61	24.19	24.19
		1.74	3.63	27.83
		1.21	2.51	30.34
		1.05	2.19	32.52
		1.04	2.17	34.70
SUA	All Codes	11.40	23.76	23.76
		1.71	3.56	27.32
		1.21	2.52	29.85
		1.06	2.20	32.05
		1.05	2.19	34.23
SWD/SUA	SWD & SUA codes	11.01	22.94	22.94
		1.71	3.55	26.50

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
SWD/SUA	SUA=504 plan codes	1.22	2.55	29.04
		1.06	2.21	31.25
		1.05	2.19	33.44
ELL/SUA	SUA & ELL Codes	9.42	19.63	19.63
		1.68	3.51	23.14
		1.29	2.70	25.83
		1.14	2.38	28.21
		1.09	2.27	30.48
		1.04	2.17	32.65
		1.03	2.14	34.79
		1.02	2.13	36.92

Table L9. Mathematics Grade 5 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL	ELL=Y	10.40	21.66	21.66
		1.84	3.83	25.49
		1.20	2.50	28.00
		1.10	2.30	30.29
		1.05	2.18	32.47
		1.02	2.13	34.60
		1.00	2.09	36.69
SWD	All Codes	11.02	22.96	22.96
		1.73	3.60	26.56
		1.16	2.42	28.97
		1.05	2.18	31.16
		1.02	2.13	33.29
		1.00	2.09	35.38
SUA	All Codes	11.13	23.19	23.19
		1.72	3.59	26.77
		1.16	2.42	29.19
		1.05	2.18	31.37
		1.02	2.12	33.49
SWD/SUA	SWD & SUA codes	10.52	21.93	21.93
		1.71	3.57	25.50

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
SWD/SUA	SUA=504 plan codes	1.17	2.44	27.94
		1.07	2.23	30.17
		1.03	2.15	32.32
		1.01	2.11	34.43
ELL/SUA	SUA & ELL Codes	8.91	18.56	18.56
		1.70	3.54	22.10
		1.25	2.61	24.71
		1.19	2.49	27.20
		1.11	2.32	29.52
		1.09	2.27	31.79
		1.04	2.17	33.96
		1.04	2.16	36.12
		1.03	2.14	38.27

Table L10. Mathematics Grade 6 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL	ELL=Y	9.79	18.14	18.14
		1.68	3.12	21.25
		1.27	2.35	23.60
		1.08	2.00	25.60
		1.07	1.98	27.58
		1.03	1.90	29.47
		1.02	1.88	31.36
		1.01	1.87	33.23
SWD	All Codes	9.35	17.32	17.32
		1.62	3.00	20.32
		1.27	2.36	22.68
		1.08	2.01	24.68
		1.05	1.94	26.62
		1.03	1.92	28.53
		1.02	1.89	30.42
		1.00	1.86	32.28
		1.00	1.85	34.13

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
SUA	All Codes	9.68	17.93	17.93
		1.63	3.02	20.95
		1.27	2.35	23.30
		1.08	2.00	25.30
		1.05	1.95	27.25
		1.03	1.90	29.15
		1.01	1.88	31.02
SWD/SUA	SUA=504 plan codes	8.87	16.43	16.43
		1.61	2.98	19.40
		1.29	2.39	21.79
		1.10	2.04	23.83
		1.06	1.96	25.79
		1.05	1.94	27.73
		1.03	1.91	29.64
		1.01	1.88	31.52
		1.01	1.87	33.39
		1.00	1.85	35.24
ELL/SUA	SUA & ELL Codes	6.65	12.31	12.31
		1.51	2.79	15.10
		1.30	2.41	17.51
		1.21	2.25	19.76
		1.18	2.18	21.94
		1.15	2.12	24.07
		1.13	2.09	26.15
		1.11	2.06	28.21
		1.10	2.04	30.25
		1.09	2.01	32.26
		1.08	2.00	34.26
		1.07	1.97	36.24
		1.05	1.94	38.18
		1.04	1.92	40.11
		1.04	1.92	42.03
		1.02	1.89	43.91
		1.01	1.87	45.78
		1.00	1.85	47.63

Table L11. Mathematics Grade 7 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL	ELL=Y	10.62	19.67	19.67
		1.43	2.65	22.33
		1.21	2.25	24.57
		1.11	2.05	26.62
		1.07	1.99	28.61
		1.04	1.93	30.54
		1.02	1.89	32.43
		1.00	1.86	34.29
SWD	All Codes	9.69	17.95	17.95
		1.42	2.63	20.58
		1.16	2.15	22.72
		1.07	1.99	24.71
		1.04	1.92	26.63
		1.03	1.92	28.55
		1.02	1.89	30.44
		1.01	1.86	32.30
SUA	All Codes	10.39	19.25	19.25
		1.43	2.64	21.89
		1.15	2.13	24.02
		1.06	1.96	25.98
		1.03	1.91	27.89
		1.03	1.91	29.80
		1.01	1.87	31.67
SWD/SUA	SUA=504 plan codes	9.26	17.14	17.14
		1.41	2.61	19.75
		1.17	2.16	21.90
		1.09	2.02	23.92
		1.06	1.96	25.88
		1.05	1.94	27.83
		1.04	1.92	29.74
		1.01	1.88	31.62
ELL/SUA	SUA & ELL Codes	6.17	11.42	11.42
		1.40	2.60	14.02
		1.27	2.35	16.37
		1.25	2.31	18.68

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/SUA	SUA & ELL Codes	1.20	2.23	20.91
		1.19	2.20	23.12
		1.16	2.15	25.27
		1.15	2.13	27.41
		1.12	2.07	29.48
		1.11	2.06	31.54
		1.10	2.03	33.57
		1.08	2.01	35.57
		1.08	1.99	37.57
		1.06	1.96	39.53
		1.05	1.95	41.47
		1.03	1.90	43.37
		1.02	1.88	45.26
		1.01	1.87	47.13
		1.00	1.86	48.98

Table L12. Mathematics Grade 8 Test Factor Analysis by Subgroup

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL	ELL=Y	9.54	17.67	17.67
		1.58	2.93	20.60
		1.26	2.33	22.93
		1.11	2.06	24.99
		1.11	2.05	27.04
		1.07	1.97	29.01
		1.03	1.92	30.93
		1.03	1.91	32.83
		1.02	1.89	34.72
SWD	All Codes	7.83	14.51	14.51
		1.45	2.69	17.19
		1.26	2.33	19.53
		1.13	2.09	21.62
		1.09	2.01	23.63
		1.07	1.98	25.61
		1.05	1.94	27.55

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
SWD	All Codes	1.03	1.90	29.45
		1.02	1.88	31.33
		1.01	1.87	33.21
SUA	All Codes	8.43	15.62	15.62
		1.47	2.72	18.34
		1.25	2.32	20.66
		1.11	2.06	22.73
		1.08	1.99	24.72
		1.06	1.97	26.69
		1.04	1.92	28.61
		1.02	1.89	30.51
		1.00	1.86	32.37
		1.00	1.85	34.22
SWD/SUA	SUA=504 plan codes	7.63	14.13	14.13
		1.45	2.69	16.82
		1.26	2.34	19.16
		1.13	2.09	21.25
		1.09	2.02	23.26
		1.08	2.00	25.26
		1.05	1.94	27.21
		1.04	1.93	29.14
		1.02	1.89	31.02
		1.01	1.88	32.90
ELL/SUA	SUA & ELL Codes	6.00	11.11	11.11
		1.41	2.61	13.72
		1.32	2.44	16.16
		1.28	2.38	18.53
		1.22	2.26	20.79
		1.19	2.21	23.01
		1.17	2.16	25.17
		1.15	2.13	27.30
		1.14	2.11	29.40
		1.13	2.08	31.49
		1.11	2.05	33.54
		1.10	2.04	35.58

Appendix L: Factor Analysis Results for Select Subgroups

Demographic Category		Extracted Factor		
		Initial	Variance Accounted for	
		Eigenvalue	%	Cumulative %
ELL/SUA	SUA & ELL Codes	1.10	2.03	37.61
		1.08	2.00	39.61
		1.06	1.96	41.57
		1.05	1.94	43.51
		1.03	1.91	45.42
		1.02	1.89	47.32
		1.02	1.88	49.20

Appendix M: Classical Test Theory Statistics

These tables support the classical test theory analyses described in Section 5, “Operational Test Data Collection and Classical Analysis.” They include item type, sample size, p -value, percent of omitted responses and the point-biserial of the key. External linking and field test items (i.e., those not contributing to students’ scores) have been omitted.

Table M1. ELA Grade 3 Classical Item Analysis

Item	Type	N-Count	p -value	% Omit	PBis Key
1	MC	175,060	.84	0.02	.38
2	MC	175,006	.59	0.05	.37
3	MC	174,903	.66	0.11	.39
4	MC	174,932	.58	0.09	.39
5	MC	174,888	.65	0.12	.44
6	MC	174,931	.72	0.10	.28
7	MC	174,883	.52	0.12	.35
8	MC	174,868	.62	0.13	.38
9	MC	174,927	.56	0.10	.42
10	MC	174,915	.46	0.10	.29
11	MC	174,861	.72	0.14	.49
12	MC	174,875	.67	0.13	.43
19	MC	174,764	.53	0.19	.37
20	MC	174,726	.69	0.21	.44
21	MC	174,713	.42	0.22	.27
22	MC	174,596	.56	0.29	.44
23	MC	174,524	.56	0.33	.30
24	MC	174,340	.52	0.43	.51
25	MC	175,049	.66	0.03	.45
26	MC	175,021	.68	0.04	.41
27	MC	174,964	.56	0.08	.31
28	MC	174,925	.48	0.10	.36
29	MC	175,014	.58	0.05	.46
30	MC	174,946	.54	0.09	.36
31	MC	174,699	.59	0.23	.39
32	CR2	174,505	.59	0.34	.56
33	CR2	174,164	.64	0.53	.63
34	CR4	173,885	.40	0.69	.69
35	CR2	174,767	.54	0.19	.56
36	CR2	173,347	.44	1.00	.62
37	CR2	173,875	.43	0.70	.61
38	CR2	173,318	.47	1.02	.59

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
39	CR2	172,966	.46	1.22	.61
40	CR4	173,126	.37	1.13	.70

Table M2. ELA Grade 4 Classical Item Analysis

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
1	MC	174,790	.80	0.02	.34
2	MC	174,782	.58	0.02	.28
3	MC	174,745	.75	0.04	.41
4	MC	174,732	.71	0.05	.48
5	MC	174,713	.70	0.06	.48
6	MC	174,672	.39	0.09	.18
7	MC	174,712	.51	0.06	.42
8	MC	174,689	.57	0.08	.45
9	MC	174,746	.70	0.04	.42
10	MC	174,701	.64	0.07	.38
11	MC	174,695	.60	0.07	.44
12	MC	174,698	.81	0.07	.39
19	MC	174,648	.68	0.10	.43
20	MC	174,610	.44	0.12	.34
21	MC	174,622	.33	0.11	.16
22	MC	174,556	.50	0.15	.28
23	MC	174,468	.40	0.20	.26
24	MC	174,348	.65	0.27	.26
25	MC	174,783	.51	0.02	.33
26	MC	174,791	.56	0.02	.37
27	MC	174,728	.36	0.05	.18
28	MC	174,720	.46	0.06	.38
29	MC	174,786	.59	0.02	.33
30	MC	174,719	.53	0.06	.36
31	MC	174,523	.68	0.17	.32
32	CR2	174,417	.66	0.23	.50
33	CR2	174,050	.58	0.44	.53
34	CR4	173,819	.44	0.57	.68
35	CR2	174,496	.61	0.19	.50
36	CR2	172,961	.45	1.06	.50
37	CR2	174,344	.59	0.27	.62
38	CR2	173,794	.57	0.59	.57
39	CR2	173,657	.57	0.67	.50

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
40	CR4	173,473	.44	0.77	.69

Table M3. ELA Grade 5 Classical Item Analysis

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
1	MC	164,559	.60	0.02	.24
2	MC	164,557	.54	0.02	.35
3	MC	164,535	.89	0.04	.41
4	MC	164,507	.44	0.05	.27
5	MC	164,501	.85	0.06	.41
6	MC	164,516	.59	0.05	.29
7	MC	164,529	.73	0.04	.39
8	MC	164,439	.55	0.10	.32
9	MC	164,460	.61	0.08	.39
10	MC	164,437	.59	0.10	.31
11	MC	164,483	.79	0.07	.41
12	MC	164,496	.42	0.06	.27
13	MC	164,496	.79	0.06	.46
14	MC	164,472	.73	0.08	.24
22	MC	164,374	.56	0.13	.33
23	MC	164,447	.67	0.09	.42
24	MC	164,409	.51	0.11	.21
25	MC	164,383	.54	0.13	.31
26	MC	164,328	.60	0.16	.40
27	MC	164,366	.62	0.14	.47
28	MC	164,343	.58	0.15	.42
29	MC	164,391	.48	0.12	.29
30	MC	164,369	.55	0.14	.27
31	MC	164,330	.43	0.16	.27
32	MC	164,355	.50	0.15	.32
33	MC	164,254	.49	0.21	.31
34	MC	164,277	.57	0.19	.37
35	MC	164,197	.44	0.24	.28
36	MC	164,561	.60	0.02	.16
37	MC	164,550	.68	0.03	.46
38	MC	164,522	.45	0.04	.25
39	MC	164,503	.44	0.06	.17
40	MC	164,557	.74	0.02	.27
41	MC	164,537	.54	0.04	.16

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
42	MC	164,396	.60	0.12	.31
43	CR2	163,876	.67	0.44	.50
44	CR2	164,072	.64	0.32	.51
45	CR4	163,742	.39	0.52	.61
46	CR2	164,035	.58	0.34	.58
47	CR2	163,967	.61	0.38	.62
48	CR2	164,245	.64	0.21	.60
49	CR2	163,983	.60	0.37	.51
50	CR2	163,538	.58	0.64	.60
51	CR4	163,671	.44	0.56	.69

Table M4. ELA Grade 6 Classical Item Analysis

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
1	MC	161,400	.82	0.01	.39
2	MC	161,317	.64	0.07	.32
3	MC	161,341	.59	0.05	.37
4	MC	161,344	.66	0.05	.36
5	MC	161,296	.53	0.08	.42
6	MC	161,263	.56	0.10	.42
7	MC	161,321	.77	0.06	.40
8	MC	161,344	.84	0.05	.39
9	MC	161,312	.53	0.07	.44
10	MC	161,247	.51	0.11	.33
11	MC	161,261	.51	0.10	.28
12	MC	161,305	.56	0.07	.42
13	MC	161,327	.43	0.06	.22
14	MC	161,304	.69	0.07	.36
15	MC	161,337	.46	0.05	.24
16	MC	161,308	.42	0.07	.30
17	MC	161,290	.65	0.08	.44
18	MC	161,278	.54	0.09	.37
19	MC	161,257	.54	0.10	.45
20	MC	161,297	.71	0.08	.41
21	MC	161,239	.59	0.11	.37
29	MC	161,231	.71	0.12	.47
30	MC	161,232	.59	0.12	.48
31	MC	161,176	.49	0.15	.40
32	MC	161,130	.57	0.18	.39

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
33	MC	161,112	.63	0.19	.39
34	MC	161,161	.53	0.16	.39
35	MC	160,985	.51	0.27	.23
36	MC	161,378	.66	0.03	.41
37	MC	161,388	.71	0.02	.31
38	MC	161,351	.69	0.05	.51
39	MC	161,300	.73	0.08	.38
40	MC	161,392	.69	0.02	.48
41	MC	161,380	.74	0.03	.37
42	MC	161,226	.72	0.12	.40
43	CR2	161,064	.71	0.22	.60
44	CR2	160,536	.69	0.55	.56
45	CR4	160,810	.59	0.38	.67
46	CR2	161,180	.71	0.15	.61
47	CR2	160,720	.66	0.44	.61
48	CR2	160,939	.72	0.30	.60
49	CR2	160,554	.70	0.54	.58
50	CR2	160,684	.71	0.46	.63
51	CR4	160,290	.50	0.70	.68

Table M5. ELA Grade 7 Classical Item Analysis

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
1	MC	152,320	.88	0.01	.40
2	MC	152,282	.68	0.04	.33
3	MC	152,255	.57	0.05	.14
4	MC	152,248	.71	0.06	.28
5	MC	152,288	.74	0.03	.32
6	MC	152,226	.62	0.07	.37
7	MC	152,260	.52	0.05	.35
8	MC	152,193	.55	0.10	.31
9	MC	152,200	.73	0.09	.22
10	MC	152,253	.89	0.06	.37
11	MC	152,162	.60	0.12	.37
12	MC	152,253	.58	0.06	.51
13	MC	152,265	.56	0.05	.39
14	MC	152,211	.65	0.08	.26
22	MC	152,132	.69	0.14	.47
23	MC	152,259	.86	0.05	.49

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
24	MC	152,198	.71	0.09	.49
25	MC	152,193	.51	0.10	.35
26	MC	152,168	.45	0.11	.40
27	MC	152,202	.68	0.09	.45
28	MC	152,150	.58	0.12	.43
29	MC	152,143	.42	0.13	.30
30	MC	152,167	.61	0.11	.47
31	MC	151,994	.41	0.23	.30
32	MC	152,040	.42	0.20	.30
33	MC	151,950	.53	0.25	.31
34	MC	152,015	.49	0.21	.39
35	MC	151,973	.63	0.24	.44
36	MC	152,280	.41	0.04	.36
37	MC	152,280	.52	0.04	.29
38	MC	152,249	.63	0.06	.38
39	MC	152,222	.51	0.08	.33
40	MC	152,301	.62	0.02	.34
41	MC	152,280	.69	0.04	.35
42	MC	152,145	.45	0.13	.18
43	CR2	151,935	.79	0.26	.56
44	CR2	151,851	.79	0.32	.58
45	CR4	151,417	.55	0.60	.68
46	CR2	151,658	.65	0.45	.57
47	CR2	151,466	.71	0.57	.65
48	CR2	152,059	.80	0.18	.59
49	CR2	151,628	.73	0.47	.58
50	CR2	151,198	.70	0.75	.62
51	CR4	151,105	.63	0.81	.67

Table M6. ELA Grade 8 Classical Item Analysis

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
1	MC	143,140	.72	0.05	.28
2	MC	143,145	.65	0.04	.15
3	MC	143,176	.79	0.02	.48
4	MC	143,144	.68	0.04	.46
5	MC	143,140	.81	0.05	.45
6	MC	143,114	.86	0.06	.33
7	MC	143,137	.78	0.05	.49

Appendix M: Classical Test Theory Statistics

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
15	MC	143,130	.77	0.05	.47
16	MC	143,091	.62	0.08	.18
17	MC	143,108	.70	0.07	.40
18	MC	143,076	.58	0.09	.41
19	MC	143,052	.75	0.11	.46
20	MC	143,079	.70	0.09	.35
21	MC	143,038	.68	0.12	.34
22	MC	142,975	.49	0.16	.35
23	MC	143,115	.77	0.06	.40
24	MC	143,076	.46	0.09	.25
25	MC	143,062	.57	0.10	.48
26	MC	143,050	.46	0.11	.37
27	MC	143,041	.44	0.12	.32
28	MC	143,033	.53	0.12	.40
29	MC	142,976	.73	0.16	.43
30	MC	143,028	.53	0.12	.27
31	MC	143,001	.64	0.14	.32
32	MC	143,013	.74	0.14	.47
33	MC	142,987	.65	0.15	.49
34	MC	142,914	.70	0.20	.50
35	MC	142,897	.48	0.22	.36
36	MC	143,189	.61	0.01	.34
37	MC	143,187	.75	0.01	.23
38	MC	143,149	.48	0.04	.15
39	MC	143,135	.68	0.05	.43
40	MC	143,180	.77	0.02	.48
41	MC	143,132	.53	0.05	.11
42	MC	143,078	.71	0.09	.46
43	CR2	142,929	.92	0.19	.51
44	CR2	142,416	.81	0.55	.59
45	CR4	142,397	.65	0.57	.67
46	CR2	143,033	.79	0.12	.55
47	CR2	142,200	.66	0.70	.53
48	CR2	142,522	.75	0.48	.58
49	CR2	142,327	.79	0.61	.59
50	CR2	141,846	.70	0.95	.58
51	CR4	141,724	.66	1.04	.69

Table M7. Mathematics Grade 3 Classical Item Analysis

Item	Type	N-Count	<i>p</i>-value	% Omit	PBis Key
1	MC	178,023	.76	0.04	.48
2	MC	177,970	.90	0.07	.41
3	MC	177,883	.85	0.11	.27
5	MC	177,720	.51	0.21	.39
6	MC	177,860	.56	0.13	.51
7	MC	177,894	.62	0.11	.53
8	MC	177,748	.67	0.19	.50
9	MC	177,840	.74	0.14	.40
10	MC	177,845	.81	0.14	.52
12	MC	177,779	.70	0.17	.46
13	MC	177,871	.66	0.12	.58
15	MC	177,736	.45	0.20	.48
16	MC	177,760	.88	0.18	.41
17	MC	177,640	.44	0.25	.46
18	MC	177,716	.77	0.21	.35
20	MC	177,725	.38	0.20	.36
21	MC	177,266	.44	0.46	.34
22	MC	176,525	.51	0.88	.62
23	MC	178,051	.94	0.02	.27
24	MC	177,866	.73	0.12	.37
25	MC	177,965	.85	0.07	.43
26	MC	177,875	.63	0.12	.51
27	MC	177,880	.91	0.12	.38
29	MC	177,927	.59	0.09	.56
30	MC	177,848	.60	0.13	.49
31	MC	177,872	.59	0.12	.53
32	MC	177,903	.76	0.10	.55
33	MC	177,821	.49	0.15	.54
35	MC	177,897	.67	0.11	.52
36	MC	177,858	.78	0.13	.42
37	MC	177,790	.60	0.17	.55
38	MC	177,773	.60	0.18	.42
39	MC	177,835	.64	0.14	.45
40	MC	177,781	.74	0.17	.52
41	MC	177,905	.76	0.10	.46
43	MC	177,583	.88	0.28	.37
44	MC	176,922	.53	0.65	.56

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
45	CR2	177,776	.49	0.17	.58
46	CR2	177,611	.65	0.27	.62
47	CR2	177,623	.58	0.26	.53
48	CR2	177,320	.49	0.43	.51
49	CR2	177,413	.60	0.38	.70
50	CR3	177,391	.60	0.39	.60
51	CR3	177,317	.30	0.43	.57
52	CR3	177,368	.48	0.40	.71

Table M8. Mathematics Grade 4 Classical Item Analysis

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
1	MC	176,627	.73	0.03	.45
2	MC	176,610	.78	0.04	.41
3	MC	176,532	.74	0.08	.47
4	MC	176,509	.59	0.10	.37
5	MC	176,441	.74	0.13	.51
6	MC	176,495	.59	0.10	.61
7	MC	176,450	.62	0.13	.54
8	MC	176,475	.61	0.12	.37
10	MC	176,374	.72	0.17	.52
11	MC	176,467	.32	0.12	.43
12	MC	176,374	.44	0.17	.55
14	MC	176,408	.72	0.15	.45
15	MC	176,469	.55	0.12	.43
16	MC	176,441	.71	0.13	.49
17	MC	176,452	.66	0.13	.48
18	MC	176,407	.78	0.15	.41
21	MC	176,087	.55	0.34	.43
22	MC	174,698	.71	1.12	.52
23	MC	176,647	.83	0.02	.49
24	MC	176,586	.81	0.05	.41
25	MC	176,542	.88	0.08	.40
26	MC	176,552	.93	0.07	.34
27	MC	176,448	.36	0.13	.36
28	MC	176,527	.76	0.09	.46
29	MC	176,555	.66	0.07	.43
30	MC	176,509	.77	0.10	.55
31	MC	176,547	.92	0.07	.36

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
32	MC	176,490	.70	0.11	.49
33	MC	176,536	.47	0.08	.38
35	MC	176,496	.59	0.10	.53
36	MC	176,561	.70	0.07	.43
37	MC	176,537	.88	0.08	.36
38	MC	176,477	.58	0.11	.45
40	MC	176,502	.66	0.10	.41
41	MC	176,543	.61	0.08	.54
43	MC	176,400	.73	0.16	.43
44	MC	176,420	.76	0.15	.54
45	MC	175,956	.41	0.41	.48
46	CR2	176,371	.43	0.17	.52
47	CR2	176,437	.52	0.14	.67
48	CR2	176,374	.55	0.17	.60
49	CR2	176,115	.48	0.32	.67
50	CR2	176,214	.32	0.26	.64
51	CR2	176,215	.78	0.26	.56
52	CR3	176,325	.53	0.20	.68
53	CR3	176,336	.46	0.19	.71
54	CR3	176,240	.52	0.25	.68
55	CR3	176,204	.58	0.27	.69

Table M9. Mathematics Grade 5 Classical Item Analysis

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
1	MC	166,523	.85	0.03	.41
2	MC	166,454	.82	0.07	.51
3	MC	166,394	.76	0.11	.51
4	MC	166,485	.75	0.05	.50
6	MC	166,435	.71	0.08	.47
7	MC	166,486	.71	0.05	.44
8	MC	166,201	.40	0.22	.35
9	MC	166,447	.74	0.07	.53
11	MC	166,419	.55	0.09	.44
12	MC	166,369	.40	0.12	.29
13	MC	166,441	.65	0.08	.40
14	MC	166,265	.49	0.18	.45
15	MC	166,388	.75	0.11	.49
16	MC	166,373	.70	0.12	.45

Appendix M: Classical Test Theory Statistics

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
17	MC	166,433	.66	0.08	.38
18	MC	166,371	.77	0.12	.48
21	MC	166,232	.63	0.20	.46
22	MC	165,813	.51	0.45	.57
23	MC	166,550	.88	0.01	.30
24	MC	166,497	.46	0.04	.52
25	MC	166,482	.59	0.05	.51
26	MC	166,479	.68	0.05	.33
27	MC	166,416	.61	0.09	.46
28	MC	166,316	.52	0.15	.55
29	MC	166,400	.59	0.10	.42
31	MC	166,480	.69	0.05	.46
32	MC	166,394	.77	0.11	.52
33	MC	166,468	.79	0.06	.50
35	MC	166,449	.73	0.07	.49
36	MC	166,480	.73	0.05	.54
37	MC	166,464	.58	0.06	.46
38	MC	166,433	.57	0.08	.53
39	MC	166,410	.43	0.10	.34
41	MC	166,442	.61	0.08	.61
42	MC	166,139	.45	0.26	.22
43	MC	166,321	.60	0.15	.44
44	MC	166,394	.64	0.11	.46
45	MC	165,900	.76	0.40	.51
46	CR2	166,011	.29	0.34	.60
47	CR2	166,330	.50	0.14	.73
48	CR2	166,143	.28	0.26	.63
49	CR2	166,148	.51	0.25	.65
50	CR2	166,148	.49	0.25	.71
51	CR2	166,168	.65	0.24	.59
52	CR3	165,756	.19	0.49	.59
53	CR3	165,675	.39	0.54	.73
54	CR3	165,855	.37	0.43	.70
55	CR3	165,520	.42	0.63	.68

Table M10. Mathematics Grade 6 Classical Item Analysis

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
1	MC	161,891	.56	0.05	.57
2	MC	161,728	.57	0.15	.48
3	MC	161,687	.35	0.17	.28
4	MC	161,626	.69	0.21	.49
6	MC	161,845	.80	0.08	.37
7	MC	161,698	.45	0.17	.54
9	MC	161,580	.24	0.24	.21
10	MC	161,804	.38	0.10	.37
11	MC	161,776	.72	0.12	.39
13	MC	161,824	.73	0.09	.49
14	MC	161,615	.65	0.22	.53
15	MC	161,841	.65	0.08	.54
16	MC	161,769	.46	0.12	.38
17	MC	161,719	.22	0.15	.33
18	MC	161,782	.53	0.12	.29
19	MC	161,741	.42	0.14	.40
20	MC	161,708	.42	0.16	.33
21	MC	161,799	.65	0.11	.52
23	MC	161,747	.64	0.14	.50
24	MC	161,598	.49	0.23	.50
25	MC	161,345	.41	0.39	.45
26	MC	160,991	.52	0.60	.45
27	MC	161,758	.78	0.13	.45
28	MC	161,934	.83	0.02	.35
29	MC	161,594	.53	0.23	.56
30	MC	161,904	.76	0.04	.37
32	MC	161,816	.72	0.10	.46
33	MC	161,818	.49	0.09	.45
34	MC	161,901	.70	0.04	.46
35	MC	161,891	.68	0.05	.33
36	MC	161,749	.54	0.14	.48
37	MC	161,886	.49	0.05	.44
38	MC	161,827	.54	0.09	.44
39	MC	161,738	.38	0.14	.24
40	MC	161,775	.82	0.12	.38
41	MC	161,839	.59	0.08	.60

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
42	MC	161,740	.45	0.14	.46
43	MC	161,642	.50	0.20	.29
44	MC	161,819	.45	0.09	.40
45	MC	161,869	.41	0.06	.48
46	MC	161,846	.85	0.08	.39
47	MC	161,744	.41	0.14	.44
50	MC	161,707	.35	0.16	.53
51	MC	161,378	.49	0.37	.52
52	CR2	161,368	.38	0.37	.64
53	CR2	161,199	.41	0.48	.65
54	CR2	161,188	.21	0.48	.62
55	CR2	161,274	.56	0.43	.65
56	CR2	161,092	.35	0.54	.70
57	CR2	161,015	.54	0.59	.69
58	CR3	161,050	.69	0.57	.62
59	CR3	160,831	.39	0.70	.67
60	CR3	160,855	.25	0.69	.64
61	CR3	156,660	.65	3.28	.57

Table M11. Mathematics Grade 7 Classical Item Analysis

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
1	MC	142,787	.76	0.09	.36
2	MC	142,813	.65	0.07	.59
3	MC	142,788	.37	0.09	.41
4	MC	142,405	.52	0.35	.43
6	MC	142,605	.40	0.21	.44
8	MC	142,590	.44	0.22	.50
9	MC	142,724	.54	0.13	.49
10	MC	142,786	.63	0.09	.56
11	MC	142,782	.54	0.09	.47
12	MC	142,614	.62	0.21	.46
13	MC	142,654	.61	0.18	.39
14	MC	142,659	.68	0.18	.42
15	MC	142,723	.57	0.13	.57
16	MC	142,738	.33	0.12	.29
18	MC	142,686	.51	0.16	.48
19	MC	142,702	.61	0.15	.62
20	MC	142,682	.49	0.16	.53

Appendix M: Classical Test Theory Statistics

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
21	MC	142,618	.30	0.20	.39
22	MC	142,786	.64	0.09	.46
23	MC	142,661	.44	0.17	.52
25	MC	142,626	.66	0.20	.50
26	MC	141,931	.74	0.69	.36
27	MC	142,800	.65	0.08	.58
28	MC	142,809	.50	0.07	.51
29	MC	142,758	.43	0.11	.27
30	MC	142,830	.70	0.06	.55
32	MC	142,661	.29	0.17	.38
33	MC	142,711	.67	0.14	.47
34	MC	142,755	.34	0.11	.33
35	MC	142,656	.40	0.18	.29
36	MC	142,700	.35	0.15	.42
37	MC	142,847	.77	0.04	.39
38	MC	142,812	.38	0.07	.44
39	MC	142,668	.48	0.17	.51
40	MC	142,759	.58	0.11	.50
41	MC	142,710	.52	0.14	.44
42	MC	142,792	.46	0.08	.42
43	MC	142,786	.64	0.09	.51
44	MC	142,695	.40	0.15	.44
45	MC	142,719	.56	0.13	.46
46	MC	142,752	.53	0.11	.48
47	MC	142,753	.47	0.11	.45
50	MC	142,669	.62	0.17	.39
51	MC	142,444	.44	0.33	.51
52	CR2	142,062	.47	0.59	.59
53	CR2	142,533	.48	0.26	.74
54	CR2	142,050	.44	0.60	.71
55	CR2	142,046	.52	0.60	.73
56	CR2	141,982	.39	0.65	.66
57	CR2	141,015	.42	1.33	.69
58	CR3	141,929	.33	0.69	.67
59	CR3	141,563	.36	0.94	.73
60	CR3	140,667	.41	1.57	.77
61	CR3	141,096	.22	1.27	.70

Table M12. Mathematics Grade 8 Classical Item Analysis

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
1	MC	107,634	.62	0.03	.37
2	MC	107,587	.46	0.08	.47
3	MC	107,493	.41	0.16	.33
4	MC	107,544	.57	0.12	.44
5	MC	107,582	.53	0.08	.36
6	MC	107,549	.66	0.11	.44
7	MC	107,490	.22	0.17	.31
8	MC	107,469	.48	0.19	.49
9	MC	107,495	.54	0.16	.48
10	MC	107,494	.49	0.16	.43
12	MC	107,574	.48	0.09	.44
14	MC	107,463	.26	0.19	.21
15	MC	107,573	.46	0.09	.50
17	MC	107,558	.51	0.10	.45
18	MC	107,428	.49	0.22	.37
19	MC	107,544	.67	0.12	.43
20	MC	107,508	.36	0.15	.39
22	MC	107,487	.32	0.17	.34
23	MC	107,462	.47	0.19	.40
24	MC	107,407	.36	0.24	.40
25	MC	107,476	.56	0.18	.39
26	MC	107,359	.58	0.29	.39
27	MC	107,529	.83	0.03	.31
28	MC	107,492	.44	0.17	.41
29	MC	107,616	.70	0.05	.37
30	MC	104,101	.48	0.46	.30
31	MC	107,569	.36	0.09	.48
32	MC	107,449	.36	0.21	.42
34	MC	107,602	.58	0.06	.35
35	MC	107,482	.54	0.17	.45
36	MC	107,539	.63	0.12	.40
37	MC	107,547	.54	0.11	.48
38	MC	107,578	.47	0.09	.39
39	MC	107,463	.58	0.19	.49
40	MC	107,525	.59	0.13	.39
42	MC	107,495	.23	0.16	.22
43	MC	107,545	.38	0.12	.32

Item	Type	N-Count	<i>p</i> -value	% Omit	PBis Key
44	MC	107,430	.39	0.22	.40
45	MC	107,554	.47	0.11	.33
46	MC	107,447	.48	0.21	.22
47	MC	107,566	.57	0.10	.48
48	MC	107,534	.63	0.13	.39
50	MC	107,511	.45	0.15	.42
51	MC	107,486	.50	0.17	.49
52	CR2	106,291	.32	1.28	.69
53	CR2	106,718	.40	0.88	.66
54	CR2	105,763	.31	1.77	.65
55	CR2	105,169	.23	2.32	.57
56	CR2	104,811	.30	2.66	.72
57	CR2	105,516	.33	2.00	.71
58	CR3	105,406	.15	2.10	.60
59	CR3	105,981	.28	1.57	.65
60	CR3	102,974	.17	4.36	.66
61	CR3	104,938	.53	2.54	.57

Appendix N: Items Flagged for DIF

These tables support the DIF information in Section 5, “Operational Test Data Collection and Classical Analysis.” They include item numbers, focal group, and directions of DIF and DIF statistics. Tables N1–N3 show items flagged by the SMD, or Mantel-Haenszel methods. No mathematics constructed-response items were flagged for DIF, so that table has been omitted. Positive values of SMD and Delta in Tables N1–N3 indicate DIF in favor of a focal group, and negative values of SMD and Delta indicate DIF against a focal group. External linking and field test items (i.e., those not contributing to students’ scores) have been omitted.

Table N1. ELA MC Item Classical DIF Flags

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
3	9	Asian	Against	1.77	916.8	-1.34
3	24	ELL	Against	1.56	534.4	-1.04
3	30	Asian	Against	1.62	707.1	-1.13
4	11	Asian	Against	1.71	723.0	-1.26
4	19	Black	Against	1.54	736.7	-1.01
5	2	Hispanic	Against	1.55	1,090.9	-1.02
5	3	Black	Against	1.54	322.6	-1.02
5	3	Hispanic	Against	1.61	467.8	-1.12
5	3	Asian	Against	1.61	163.7	-1.12
5	3	ELL	Against	1.85	757.4	-1.44
5	10	Hispanic	Against	1.53	1,061.6	-1.00
5	10	CBT	In Favor	0.60	134.7	1.21
5	37	ELL	Against	1.64	560.4	-1.16
6	1	Asian	Against	1.80	512.8	-1.38
6	1	ELL	Against	3.10	2,698.9	-2.66
6	3	Female	Against	1.78	2,637.6	-1.36
6	3	Black	Against	1.95	1,977.7	-1.57
6	3	Hispanic	Against	1.77	1,745.5	-1.34
6	3	High Needs	Against	2.08	3,289.9	-1.72
6	3	CBT	In Favor	0.62	105.9	1.13
6	4	CBT	In Favor	0.61	103.2	1.18
6	5	CBT	In Favor	0.65	87.9	1.01
6	6	Female	Against	1.64	1,905.1	-1.17
6	8	ELL	Against	1.60	448.5	-1.10
6	9	CBT	In Favor	0.63	94.3	1.07
6	30	Female	Against	1.59	1,540.3	-1.10
6	30	Black	Against	1.65	982.6	-1.17
6	38	Black	Against	1.66	854.9	-1.19
6	38	Hispanic	Against	1.82	1,480.5	-1.41

Appendix N: Items Flagged for DIF

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
6	38	Asian	Against	2.00	841.9	-1.62
6	38	High Needs	Against	1.85	1,704.9	-1.44
6	38	ELL	Against	2.04	983.5	-1.67
6	38	CBT	In Favor	0.51	145.2	1.57
7	1	Asian	Against	1.72	235.7	-1.28
7	1	ELL	Against	2.24	1,143.9	-1.89
7	5	Female	Against	1.85	2,340.6	-1.44
7	5	Hispanic	Against	1.90	1,784.3	-1.51
7	5	Asian	Against	1.84	722.7	-1.43
7	5	High Needs	Against	1.93	2,021.2	-1.55
7	5	ELL	Against	2.14	1,243.8	-1.79
7	5	CBT	In Favor	0.64	94.3	1.05
7	6	Black	Against	1.54	759.8	-1.02
7	10	Black	Against	1.70	463.6	-1.25
7	10	Hispanic	Against	2.12	1,126.8	-1.76
7	10	Asian	Against	2.12	459.5	-1.76
7	10	High Needs	Against	1.97	920.3	-1.59
7	10	ELL	Against	1.96	769.7	-1.58
7	10	CBT	In Favor	0.50	94.3	1.63
7	12	Black	Against	1.55	695.6	-1.03
7	12	Hispanic	Against	1.70	1,221.2	-1.24
7	12	High Needs	Against	1.69	1,426.3	-1.23
7	12	ELL	Against	1.56	304.9	-1.05
7	23	ELL	Against	1.68	501.3	-1.22
7	23	CBT	In Favor	0.58	70.3	1.29
7	29	Asian	Against	1.58	603.7	-1.08
7	30	ELL	Against	1.58	362.8	-1.08
7	31	Asian	Against	1.56	546.2	-1.05
7	39	Black	Against	1.69	1,179.1	-1.24
7	39	Hispanic	Against	1.92	2,241.7	-1.53
7	39	Asian	Against	2.32	1,983.4	-1.98
7	39	High Needs	Against	1.85	2,502.8	-1.44
7	39	ELL	Against	1.74	515.8	-1.31
7	40	Asian	In Favor	0.64	471.1	1.05
8	1	Asian	Against	1.71	630.0	-1.26
8	1	ELL	Against	1.75	593.6	-1.32
8	3	Hispanic	Against	1.61	654.5	-1.12
8	3	Asian	Against	1.92	530.7	-1.54

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
8	3	ELL	Against	2.86	1,854.8	-2.47
8	20	ELL	Against	1.60	400.1	-1.10
8	21	Hispanic	Against	1.53	792.4	-1.00
8	21	Asian	Against	1.65	550.0	-1.18
8	42	ELL	Against	1.76	523.8	-1.33

Table N2. ELA CR Item Classical DIF Flags

Grade	Item	Subgroup	DIF	SMD	Effect
3	34	CBT	Against	0.0447	0.044
3	37	CBT	Against	0.0058	0.009
3	38	Asian	In Favor	0.1106	0.188
3	39	Asian	In Favor	0.1218	0.176
4	33	CBT	Against	0.0647	0.102
4	34	CBT	Against	0.0693	0.065
4	39	High Needs	In Favor	0.1082	0.174
4	40	CBT	Against	0.0988	0.089
5	43	Hispanic	In Favor	0.1168	0.181
5	43	Asian	In Favor	0.1138	0.177
5	43	High Needs	In Favor	0.1447	0.224
5	44	CBT	Against	0.0283	0.045
5	45	Asian	In Favor	0.2158	0.205
5	45	CBT	Against	0.1207	0.117
5	51	Asian	In Favor	0.2256	0.205
5	51	CBT	Against	0.0750	0.069
6	43	High Needs	In Favor	0.1225	0.193
6	44	Hispanic	In Favor	0.1322	0.208
6	44	Asian	In Favor	0.1243	0.198
6	44	High Needs	In Favor	0.1772	0.279
6	45	Female	In Favor	0.1774	0.176
6	45	CBT	Against	0.1774	0.176
6	47	CBT	Against	0.0810	0.117
6	48	Black	In Favor	0.1462	0.225
6	48	Hispanic	In Favor	0.1250	0.192
6	48	Asian	In Favor	0.1087	0.173
6	48	High Needs	In Favor	0.1575	0.242
6	49	Hispanic	In Favor	0.1330	0.206
6	49	High Needs	In Favor	0.1498	0.232
6	49	ELL	In Favor	0.1599	0.247

Grade	Item	Subgroup	DIF	SMD	Effect
6	51	Asian	In Favor	0.1963	0.176
6	51	CBT	Against	0.1466	0.133
7	45	Female	In Favor	0.2646	0.245
7	45	CBT	Against	0.2646	0.245
7	47	Female	In Favor	0.1172	0.174
7	47	Black	In Favor	0.1489	0.219
7	47	Hispanic	In Favor	0.1301	0.191
7	47	High Needs	In Favor	0.1446	0.215
7	48	Black	In Favor	0.1147	0.203
7	48	Hispanic	In Favor	0.1226	0.218
7	48	High Needs	In Favor	0.1355	0.245
7	49	High Needs	In Favor	0.1312	0.193
7	50	Black	In Favor	0.1324	0.197
7	50	Hispanic	In Favor	0.1474	0.221
7	50	Asian	In Favor	0.1409	0.216
7	50	High Needs	In Favor	0.1814	0.274
7	50	ELL	In Favor	0.1191	0.180
7	51	Female	In Favor	0.2174	0.207
7	51	Asian	In Favor	0.1826	0.172
7	51	CBT	Against	0.2174	0.207
8	43	ELL	Against	-0.1579	0.385
8	46	High Needs	In Favor	0.1072	0.184
8	47	High Needs	In Favor	0.1651	0.234
8	48	High Needs	In Favor	0.1112	0.177
8	49	Black	In Favor	0.1023	0.173
8	49	Hispanic	In Favor	0.1118	0.191
8	49	High Needs	In Favor	0.1083	0.185
8	49	ELL	In Favor	0.1027	0.177
8	51	Female	In Favor	0.2151	0.210
8	51	CBT	Against	0.2151	0.210

Table N3. Mathematics MC Item Classical DIF Flags

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
3	2	Asian	In Favor	0.62	109.0	1.11
3	3	CBT	Against	2.00	212.1	-1.62
3	6	CBT	In Favor	0.57	139.0	1.32
3	7	Black	Against	1.56	754.8	-1.05
3	7	CBT	In Favor	0.65	72.4	1.01

Grade	Item	Subgroup	DIF	Alpha	MH	Delta
3	13	Black	Against	1.57	672.7	-1.06
3	13	Hispanic	Against	1.71	1,285.9	-1.27
3	13	ELL	Against	1.53	595.0	-1.01
3	16	CBT	In Favor	0.43	88.6	1.99
3	31	CBT	In Favor	0.57	127.7	1.31
3	43	CBT	Against	1.57	57.1	-1.06
4	23	Asian	In Favor	0.58	228.9	1.27
5	15	CBT	Against	1.56	55.5	-1.05
5	23	ELL	Against	1.70	585.8	-1.25
6	1	Female	Against	1.85	2,423.0	-1.44
6	6	Asian	Against	1.67	376.2	-1.20
6	6	ELL	Against	2.01	1,293.4	-1.64
6	28	Black	In Favor	0.65	424.1	1.02
6	28	ELL	Against	1.88	889.5	-1.49
6	30	ELL	Against	1.62	574.6	-1.13
7	6	Female	Against	1.54	1,254.7	-1.01
7	22	ELL	Against	1.64	501.9	-1.17
7	37	ELL	Against	1.78	793.4	-1.36
8	15	Female	Against	1.54	920.5	-1.02
8	17	Female	Against	1.60	1,134.7	-1.10
8	17	Asian	Against	1.76	442.9	-1.33
8	17	CBT	In Favor	0.60	50.5	1.21
8	44	CBT	In Favor	0.63	44.2	1.09

Table N4. Mathematics CR Item Classical DIF Flags

Grade	Item	Subgroup	DIF	SMD	Effect
3	50	CBT	Against	0.1054	0.098
4	46	CBT	In Favor	0.0020	0.002
4	49	CBT	Against	0.0702	0.083
4	51	ELL	Against	-0.1481	0.210
6	61	ELL	Against	-0.2138	0.186
7	52	CBT	Against	0.1050	0.124
8	55	CBT	In Favor	0.0519	0.072

Appendix O: IRT Statistics

External linking and field test items (i.e., those not contributing to students' scores) have been omitted.

Table O1. ELA Grade 3 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	279.65	8	67.91	466.33	Y
2	3PL	283.64	8	68.91	466.18	Y
3	3PL	249.15	8	60.29	465.91	Y
4	3PL	467.21	8	114.80	465.98	Y
5	3PL	383.30	8	93.82	465.87	Y
6	3PL	462.07	8	113.52	465.98	Y
7	3PL	216.37	8	52.09	465.85	Y
8	3PL	286.51	8	69.63	465.81	Y
9	3PL	452.12	8	111.03	465.97	Y
10	3PL	190.20	8	45.55	465.94	Y
11	3PL	621.24	8	153.31	465.79	Y
12	3PL	709.83	8	175.46	465.83	Y
19	3PL	367.78	8	89.94	465.54	Y
20	3PL	265.13	8	64.28	465.43	Y
21	3PL	263.83	8	63.96	465.40	Y
22	3PL	440.89	8	108.22	465.09	Y
23	3PL	221.69	8	53.42	464.90	Y
24	3PL	935.82	8	231.96	464.41	Y
25	3PL	298.54	8	72.64	466.30	Y
26	3PL	278.96	8	67.74	466.22	Y
27	3PL	305.27	8	74.32	466.07	Y
28	3PL	487.28	8	119.82	465.97	Y
29	3PL	574.85	8	141.71	466.20	Y
30	3PL	339.78	8	82.95	466.02	Y
31	3PL	417.07	8	102.27	465.36	Y
32	2PPC	800.39	17	134.35	464.85	Y
33	2PPC	849.07	17	142.70	463.94	Y
34	2PPC	691.97	35	78.52	463.19	Y
35	2PPC	698.54	17	116.88	465.54	Y
36	2PPC	861.11	17	144.76	461.76	Y
37	2PPC	1268.2	17	214.57	463.17	Y
38	2PPC	794.31	17	133.31	461.68	Y
39	2PPC	1177.4	17	199.01	460.74	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
40	2PPC	615.93	35	69.43	461.17	Y

Table O2. ELA Grade 4 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	303.00	8	73.75	465.91	Y
2	3PL	253.67	8	61.42	465.89	Y
3	3PL	341.49	8	83.37	465.79	Y
4	3PL	493.83	8	121.46	465.75	Y
5	3PL	537.76	8	132.44	465.70	Y
6	3PL	227.96	8	54.99	465.59	Y
7	3PL	486.65	8	119.66	465.70	Y
8	3PL	568.05	8	140.01	465.64	Y
9	3PL	410.53	8	100.63	465.79	Y
10	3PL	305.61	8	74.40	465.67	Y
11	3PL	451.50	8	110.87	465.66	Y
12	3PL	431.97	8	105.99	465.66	Y
19	3PL	376.87	8	92.22	465.53	Y
20	3PL	482.02	8	118.51	465.43	Y
21	3PL	142.66	8	33.66	465.46	Y
22	3PL	258.09	8	62.52	465.29	Y
23	3PL	438.20	8	107.55	465.05	Y
24	3PL	1690.5	8	420.62	464.73	Y
25	3PL	307.73	8	74.93	465.89	Y
26	3PL	314.83	8	76.71	465.91	Y
27	3PL	99.40	8	22.85	465.74	Y
28	3PL	456.85	8	112.21	465.72	Y
29	3PL	266.15	8	64.54	465.90	Y
30	3PL	336.38	8	82.10	465.72	Y
31	3PL	255.85	8	61.96	465.20	Y
32	2PPC	694.20	17	116.14	464.91	Y
33	2PPC	1345.8	17	227.89	463.94	Y
34	2PPC	642.74	35	72.64	463.32	Y
35	2PPC	973.71	17	164.07	465.13	Y
36	2PPC	594.03	17	98.96	461.03	Y
37	2PPC	555.39	17	92.33	464.72	Y
38	2PPC	1100.8	17	185.87	463.25	Y
39	2PPC	623.59	17	104.03	462.89	Y
40	2PPC	793.13	35	90.61	462.40	Y

Table O3. ELA Grade 5 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	187.36	8	44.84	438.64	Y
2	3PL	228.62	8	55.15	438.63	Y
3	3PL	343.29	8	83.82	438.57	Y
4	3PL	339.72	8	82.93	438.50	Y
5	3PL	315.68	8	76.92	438.48	Y
6	3PL	415.98	8	101.99	438.52	Y
7	3PL	272.18	8	66.05	438.56	Y
8	3PL	212.12	8	51.03	438.32	Y
9	3PL	271.61	8	65.90	438.37	Y
10	3PL	251.34	8	60.84	438.31	Y
11	3PL	350.57	8	85.64	438.43	Y
12	3PL	367.70	8	89.92	438.47	Y
13	3PL	303.03	8	73.76	438.47	Y
14	3PL	527.40	8	129.85	438.41	Y
22	3PL	392.61	8	96.15	438.14	Y
23	3PL	291.79	8	70.95	438.34	Y
24	3PL	177.02	8	42.26	438.24	Y
25	3PL	439.06	8	107.76	438.17	Y
26	3PL	283.23	8	68.81	438.02	Y
27	3PL	539.82	8	132.96	438.12	Y
28	3PL	386.69	8	94.67	438.06	Y
29	3PL	198.80	8	47.70	438.19	Y
30	3PL	146.84	8	34.71	438.13	Y
31	3PL	266.42	8	64.61	438.03	Y
32	3PL	245.35	8	59.34	438.09	Y
33	3PL	292.98	8	71.24	437.82	Y
34	3PL	613.28	8	151.32	437.89	Y
35	3PL	275.78	8	66.95	437.67	Y
36	3PL	118.94	8	27.73	438.64	Y
37	3PL	339.22	8	82.81	438.61	Y
38	3PL	280.30	8	68.07	438.54	Y
39	3PL	97.58	8	22.39	438.49	Y
40	3PL	156.37	8	37.09	438.63	Y
41	3PL	502.20	8	123.55	438.58	Y
42	3PL	249.23	8	60.31	438.20	Y
43	2PPC	578.63	17	96.32	436.82	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
44	2PPC	390.41	17	64.04	437.34	Y
45	2PPC	536.15	35	59.90	436.46	Y
46	2PPC	659.74	17	110.23	437.24	Y
47	2PPC	672.73	17	112.46	437.06	Y
48	2PPC	1137.1	17	192.10	437.80	Y
49	2PPC	496.03	17	82.15	437.10	Y
50	2PPC	353.04	17	57.63	435.91	Y
51	2PPC	662.44	35	74.99	436.27	Y

Table O4. ELA Grade 6 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	215.41	8	51.85	430.40	Y
2	3PL	201.53	8	48.38	430.18	Y
3	3PL	272.78	8	66.19	430.24	Y
4	3PL	219.17	8	52.79	430.25	Y
5	3PL	341.71	8	83.43	430.12	Y
6	3PL	301.58	8	73.40	430.03	Y
7	3PL	491.51	8	120.88	430.19	Y
8	3PL	155.27	8	36.82	430.25	Y
9	3PL	381.02	8	93.26	430.17	Y
10	3PL	297.97	8	72.49	429.99	Y
11	3PL	131.03	8	30.76	430.03	Y
12	3PL	322.44	8	78.61	430.15	Y
13	3PL	95.59	8	21.90	430.21	Y
14	3PL	200.39	8	48.10	430.14	Y
15	3PL	164.14	8	39.04	430.23	Y
16	3PL	482.97	8	118.74	430.15	Y
17	3PL	330.00	8	80.50	430.11	Y
18	3PL	307.88	8	74.97	430.07	Y
19	3PL	384.07	8	94.02	430.02	Y
20	3PL	386.92	8	94.73	430.13	Y
21	3PL	445.77	8	109.44	429.97	Y
29	3PL	246.01	8	59.50	429.95	Y
30	3PL	323.72	8	78.93	429.95	Y
31	3PL	422.14	8	103.54	429.80	Y
32	3PL	345.74	8	84.44	429.68	Y
33	3PL	193.75	8	46.44	429.63	Y
34	3PL	289.93	8	70.48	429.76	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
35	3PL	325.35	8	79.34	429.29	Y
36	3PL	248.55	8	60.14	430.34	Y
37	3PL	206.07	8	49.52	430.37	Y
38	3PL	258.44	8	62.61	430.27	Y
39	3PL	194.69	8	46.67	430.13	Y
40	3PL	231.30	8	55.83	430.38	Y
41	3PL	381.36	8	93.34	430.35	Y
42	3PL	164.11	8	39.03	429.94	Y
43	2PPC	416.87	17	68.58	429.50	Y
44	2PPC	539.74	17	89.65	428.10	Y
45	2PPC	567.50	35	63.65	428.83	Y
46	2PPC	508.64	17	84.32	429.81	Y
47	2PPC	545.80	17	90.69	428.59	Y
48	2PPC	447.95	17	73.91	429.17	Y
49	2PPC	577.22	17	96.08	428.14	Y
50	2PPC	804.60	17	135.07	428.49	Y
51	2PPC	631.37	35	71.28	427.44	Y

Table O5. ELA Grade 7 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	170.51	8	40.63	406.19	Y
2	3PL	181.06	8	43.26	406.09	Y
3	3PL	699.09	8	172.77	406.01	Y
4	3PL	400.55	8	98.14	405.99	Y
5	3PL	535.80	8	131.95	406.10	Y
6	3PL	174.32	8	41.58	405.94	Y
7	3PL	292.04	8	71.01	406.03	Y
8	3PL	187.62	8	44.90	405.85	Y
9	3PL	375.15	8	91.79	405.87	Y
10	3PL	227.58	8	54.89	406.01	Y
11	3PL	226.73	8	54.68	405.77	Y
12	3PL	440.01	8	108.00	406.01	Y
13	3PL	334.03	8	81.51	406.04	Y
14	3PL	465.83	8	114.46	405.90	Y
22	3PL	264.64	8	64.16	405.69	Y
23	3PL	230.08	8	55.52	406.02	Y
24	3PL	273.99	8	66.50	405.86	Y
25	3PL	291.28	8	70.82	405.85	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
26	3PL	577.95	8	142.49	405.78	Y
27	3PL	224.77	8	54.19	405.87	Y
28	3PL	301.89	8	73.47	405.73	Y
29	3PL	337.16	8	82.29	405.71	Y
30	3PL	344.97	8	84.24	405.78	Y
31	3PL	454.26	8	111.56	405.32	Y
32	3PL	425.17	8	104.29	405.44	Y
33	3PL	168.65	8	40.16	405.20	Y
34	3PL	370.70	8	90.67	405.37	Y
35	3PL	308.88	8	75.22	405.26	Y
36	3PL	526.52	8	129.63	406.08	Y
37	3PL	251.71	8	60.93	406.08	Y
38	3PL	194.55	8	46.64	406.00	Y
39	3PL	299.50	8	72.88	405.93	Y
40	3PL	191.56	8	45.89	406.14	Y
41	3PL	252.92	8	61.23	406.08	Y
42	3PL	295.81	8	71.95	405.72	Y
43	2PPC	276.42	17	44.49	405.16	Y
44	2PPC	453.44	17	74.85	404.94	Y
45	2PPC	536.91	35	59.99	403.78	Y
46	2PPC	867.31	17	145.83	404.42	Y
47	2PPC	1031.5	17	173.98	403.91	Y
48	2PPC	356.35	17	58.20	405.49	Y
49	2PPC	623.12	17	103.95	404.34	Y
50	2PPC	369.21	17	60.40	403.19	Y
51	2PPC	538.73	35	60.21	402.95	Y

Table O6. ELA Grade 8 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	122.76	8	28.69	381.71	Y
2	3PL	136.87	8	32.22	381.72	Y
3	3PL	202.29	8	48.57	381.80	Y
4	3PL	309.18	8	75.29	381.72	Y
5	3PL	199.89	8	47.97	381.71	Y
6	3PL	338.11	8	82.53	381.64	Y
7	3PL	210.25	8	50.56	381.70	Y
15	3PL	205.16	8	49.29	381.68	Y
16	3PL	685.37	8	169.34	381.58	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
17	3PL	198.64	8	47.66	381.62	Y
18	3PL	324.41	8	79.10	381.54	Y
19	3PL	209.69	8	50.42	381.47	Y
20	3PL	171.71	8	40.93	381.54	Y
21	3PL	161.96	8	38.49	381.43	Y
22	3PL	385.05	8	94.26	381.27	Y
23	3PL	174.08	8	41.52	381.64	Y
24	3PL	265.69	8	64.42	381.54	Y
25	3PL	549.62	8	135.40	381.50	Y
26	3PL	827.39	8	204.85	381.47	Y
27	3PL	559.61	8	137.90	381.44	Y
28	3PL	555.56	8	136.89	381.42	Y
29	3PL	252.42	8	61.10	381.27	Y
30	3PL	235.84	8	56.96	381.41	Y
31	3PL	244.65	8	59.16	381.34	Y
32	3PL	253.48	8	61.37	381.37	Y
33	3PL	384.90	8	94.23	381.30	Y
34	3PL	351.72	8	85.93	381.10	Y
35	3PL	529.65	8	130.41	381.06	Y
36	3PL	197.94	8	47.49	381.84	Y
37	3PL	1105.6	8	274.40	381.83	Y
38	3PL	193.48	8	46.37	381.73	Y
39	3PL	223.48	8	53.87	381.69	Y
40	3PL	222.63	8	53.66	381.81	Y
41	3PL	118.84	8	27.71	381.69	Y
42	3PL	290.62	8	70.65	381.54	Y
43	2PPC	202.20	17	31.76	381.14	Y
44	2PPC	335.64	17	54.65	379.78	Y
45	2PPC	825.16	35	94.44	379.73	Y
46	2PPC	171.45	17	26.49	381.42	Y
47	2PPC	791.70	17	132.86	379.20	Y
48	2PPC	283.54	17	45.71	380.06	Y
49	2PPC	267.41	17	42.95	379.54	Y
50	2PPC	670.57	17	112.09	378.26	Y
51	2PPC	790.04	35	90.24	377.93	Y

Table O7. Mathematics Grade 3 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	227.91	8	54.98	474.73	Y
2	3PL	454.16	8	111.54	474.59	Y
3	3PL	459.79	8	112.95	474.35	Y
5	3PL	624.15	8	154.04	473.92	Y
6	3PL	343.70	8	83.93	474.29	Y
7	3PL	276.04	8	67.01	474.38	Y
8	3PL	267.81	8	64.95	473.99	Y
9	3PL	324.25	8	79.06	474.24	Y
10	3PL	573.12	8	141.28	474.25	Y
12	3PL	209.17	8	50.29	474.08	Y
13	3PL	314.92	8	76.73	474.32	Y
15	3PL	604.99	8	149.25	473.96	Y
16	3PL	591.54	8	145.88	474.03	Y
17	3PL	530.92	8	130.73	473.71	Y
18	3PL	732.74	8	181.18	473.91	Y
20	3PL	569.93	8	140.48	473.93	Y
21	3PL	413.18	8	101.29	472.71	Y
22	3PL	647.54	8	159.89	470.73	Y
23	3PL	147.12	8	34.78	474.80	Y
24	3PL	203.45	8	48.86	474.31	Y
25	3PL	439.52	8	107.88	474.57	Y
26	3PL	255.94	8	61.99	474.33	Y
27	3PL	498.66	8	122.66	474.35	Y
29	3PL	492.97	8	121.24	474.47	Y
30	3PL	287.67	8	69.92	474.26	Y
31	3PL	395.17	8	96.79	474.33	Y
32	3PL	277.94	8	67.49	474.41	Y
33	3PL	433.10	8	106.27	474.19	Y
35	3PL	224.91	8	54.23	474.39	Y
36	3PL	302.50	8	73.62	474.29	Y
37	3PL	322.02	8	78.51	474.11	Y
38	3PL	456.74	8	112.19	474.06	Y
39	3PL	746.07	8	184.52	474.23	Y
40	3PL	306.51	8	74.63	474.08	Y
41	3PL	261.90	8	63.47	474.41	Y
43	3PL	215.40	8	51.85	473.55	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
44	3PL	746.49	8	184.62	471.79	Y
45	2PPC	425.86	17	70.12	474.07	Y
46	2PPC	1000.0	17	168.59	473.63	Y
47	2PPC	2128.9	17	362.19	473.66	Y
48	2PPC	479.95	17	79.40	472.85	Y
49	2PPC	285.95	17	46.12	473.10	Y
50	2PPC	585.36	26	77.57	473.04	Y
51	2PPC	1023.1	26	138.27	472.85	Y
52	2PPC	598.28	26	79.36	472.98	Y

Table O8. Mathematics Grade 4 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	163.88	8	38.97	471.01	Y
2	3PL	134.82	8	31.70	470.96	Y
3	3PL	213.90	8	51.47	470.75	Y
4	3PL	267.30	8	64.83	470.69	Y
5	3PL	183.09	8	43.77	470.51	Y
6	3PL	291.87	8	70.97	470.65	Y
7	3PL	308.43	8	75.11	470.53	Y
8	3PL	735.97	8	181.99	470.60	Y
10	3PL	207.81	8	49.95	470.33	Y
11	3PL	930.93	8	230.73	470.58	Y
12	3PL	800.67	8	198.17	470.33	Y
14	3PL	180.42	8	43.11	470.42	Y
15	3PL	188.78	8	45.20	470.58	Y
16	3PL	210.03	8	50.51	470.51	Y
17	3PL	237.70	8	57.42	470.54	Y
18	3PL	303.51	8	73.88	470.42	Y
21	3PL	227.56	8	54.89	469.57	Y
22	3PL	200.15	8	48.04	465.86	Y
23	3PL	526.41	8	129.60	471.06	Y
24	3PL	230.57	8	55.64	470.90	Y
25	3PL	779.62	8	192.90	470.78	Y
26	3PL	555.45	8	136.86	470.81	Y
27	3PL	392.57	8	96.14	470.53	Y
28	3PL	191.56	8	45.89	470.74	Y
29	3PL	195.48	8	46.87	470.81	Y
30	3PL	308.86	8	75.21	470.69	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
31	3PL	464.06	8	114.02	470.79	Y
32	3PL	267.46	8	64.87	470.64	Y
33	3PL	237.55	8	57.39	470.76	Y
35	3PL	264.58	8	64.14	470.66	Y
36	3PL	251.72	8	60.93	470.83	Y
37	3PL	250.31	8	60.58	470.77	Y
38	3PL	168.47	8	40.12	470.61	Y
40	3PL	176.29	8	42.07	470.67	Y
41	3PL	262.21	8	63.55	470.78	Y
43	3PL	291.66	8	70.91	470.40	Y
44	3PL	288.30	8	70.07	470.45	Y
45	3PL	1076.1	8	267.01	469.22	Y
46	2PPC	2278.0	17	387.75	470.32	Y
47	2PPC	279.79	17	45.07	470.50	Y
48	2PPC	244.53	17	39.02	470.33	Y
49	2PPC	128.53	17	19.13	469.64	Y
50	2PPC	506.69	17	83.98	469.90	Y
51	2PPC	154.59	17	23.60	469.91	Y
52	2PPC	1019.1	26	137.72	470.20	Y
53	2PPC	538.20	26	71.03	470.23	Y
54	2PPC	768.10	26	102.91	469.97	Y
55	2PPC	685.64	26	91.48	469.88	Y

Table O9. Mathematics Grade 5 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	682.91	8	168.73	443.73	Y
2	3PL	383.61	8	93.90	443.55	Y
3	3PL	389.73	8	95.43	443.39	Y
4	3PL	429.35	8	105.34	443.63	Y
6	3PL	128.79	8	30.20	443.50	Y
7	3PL	359.65	8	87.91	443.63	Y
8	3PL	222.30	8	53.58	442.87	Y
9	3PL	782.30	8	193.58	443.53	Y
11	3PL	638.47	8	157.62	443.46	Y
12	3PL	196.01	8	47.00	443.32	Y
13	3PL	337.73	8	82.43	443.51	Y
14	3PL	1066.2	8	264.54	443.05	Y
15	3PL	209.67	8	50.42	443.37	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
16	3PL	204.66	8	49.16	443.33	Y
17	3PL	326.44	8	79.61	443.49	Y
18	3PL	370.20	8	90.55	443.33	Y
21	3PL	412.37	8	101.09	442.96	Y
22	3PL	221.06	8	53.26	441.84	Y
23	3PL	547.80	8	134.95	443.81	Y
24	3PL	221.78	8	53.45	443.66	Y
25	3PL	152.22	8	36.06	443.62	Y
26	3PL	153.03	8	36.26	443.62	Y
27	3PL	140.74	8	33.19	443.45	Y
28	3PL	282.35	8	68.59	443.18	Y
29	3PL	218.10	8	52.53	443.41	Y
31	3PL	166.93	8	39.73	443.62	Y
32	3PL	158.07	8	37.52	443.39	Y
33	3PL	386.43	8	94.61	443.59	Y
35	3PL	424.89	8	104.22	443.54	Y
36	3PL	287.14	8	69.79	443.62	Y
37	3PL	412.92	8	101.23	443.58	Y
38	3PL	216.26	8	52.07	443.49	Y
39	3PL	147.83	8	34.96	443.43	Y
41	3PL	229.44	8	55.36	443.52	Y
42	3PL	371.12	8	90.78	442.71	Y
43	3PL	184.62	8	44.16	443.19	Y
44	3PL	194.44	8	46.61	443.39	Y
45	3PL	404.10	8	99.02	442.07	Y
46	2PPC	432.81	17	71.31	442.37	Y
47	2PPC	621.17	17	103.61	443.22	Y
48	2PPC	255.98	17	40.98	442.72	Y
49	2PPC	671.35	17	112.22	442.73	Y
50	2PPC	109.45	17	15.85	442.73	Y
51	2PPC	1021.5	17	172.27	442.79	Y
52	2PPC	151.34	26	17.38	441.69	Y
53	2PPC	771.04	26	103.32	441.47	Y
54	2PPC	345.34	26	44.28	441.95	Y
55	2PPC	574.52	26	76.07	441.06	Y

Table O10. Mathematics Grade 6 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	235.23	8	56.81	430.96	Y
2	3PL	116.86	8	27.22	430.53	Y
3	3PL	107.16	8	24.79	430.42	Y
4	3PL	182.90	8	43.73	430.26	Y
6	3PL	359.80	8	87.95	430.84	Y
7	3PL	368.08	8	90.02	430.45	Y
9	3PL	510.01	8	125.50	430.13	Y
10	3PL	241.23	8	58.31	430.73	Y
11	3PL	168.05	8	40.01	430.66	Y
13	3PL	215.34	8	51.83	430.78	Y
14	3PL	136.70	8	32.18	430.23	Y
15	3PL	195.95	8	46.99	430.83	Y
16	3PL	192.51	8	46.13	430.64	Y
17	3PL	245.53	8	59.38	430.50	Y
18	3PL	237.36	8	57.34	430.67	Y
19	3PL	160.71	8	38.18	430.56	Y
20	3PL	162.55	8	38.64	430.47	Y
21	3PL	231.71	8	55.93	430.72	Y
23	3PL	298.08	8	72.52	430.58	Y
24	3PL	156.50	8	37.13	430.18	Y
25	3PL	177.62	8	42.41	429.51	Y
26	3PL	119.34	8	27.84	428.57	Y
27	3PL	408.18	8	100.04	430.61	Y
28	3PL	1179.8	8	292.96	431.08	Y
29	3PL	329.12	8	80.28	430.17	Y
30	3PL	1015.3	8	251.84	431.00	Y
32	3PL	217.80	8	52.45	430.76	Y
33	3PL	197.46	8	47.36	430.77	Y
34	3PL	349.52	8	85.38	430.99	Y
35	3PL	200.44	8	48.11	430.96	Y
36	3PL	195.10	8	46.78	430.58	Y
37	3PL	137.68	8	32.42	430.95	Y
38	3PL	178.01	8	42.50	430.79	Y
39	3PL	296.16	8	72.04	430.55	Y
40	3PL	252.89	8	61.22	430.65	Y
41	3PL	212.78	8	51.19	430.82	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
42	3PL	145.21	8	34.30	430.56	Y
43	3PL	302.47	8	73.62	430.30	Y
44	3PL	114.17	8	26.54	430.77	Y
45	3PL	291.66	8	70.92	430.90	Y
46	3PL	278.74	8	67.69	430.84	Y
47	3PL	199.26	8	47.81	430.57	Y
50	3PL	324.02	8	79.01	430.47	Y
51	3PL	220.62	8	53.16	429.60	Y
52	2PPC	964.18	17	162.44	429.57	Y
53	2PPC	315.54	17	51.20	429.12	Y
54	2PPC	1141.8	17	192.90	429.09	Y
55	2PPC	378.09	17	61.93	429.32	Y
56	2PPC	219.05	17	34.65	428.83	Y
57	2PPC	204.85	17	32.22	428.63	Y
58	2PPC	2050.4	26	280.74	428.72	Y
59	2PPC	311.77	26	39.63	428.14	Y
60	2PPC	467.96	26	61.29	428.20	Y
61	2PPC	158.94	26	18.44	417.02	Y

Table O11. Mathematics Grade 7 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	565.87	8	139.47	379.84	Y
2	3PL	401.59	8	98.40	379.91	Y
3	3PL	153.47	8	36.37	379.84	Y
4	3PL	190.71	8	45.68	378.82	Y
6	3PL	111.68	8	25.92	379.35	Y
8	3PL	257.43	8	62.36	379.31	Y
9	3PL	356.78	8	87.19	379.67	Y
10	3PL	367.87	8	89.97	379.83	Y
11	3PL	93.32	8	21.33	379.82	Y
12	3PL	145.86	8	34.47	379.38	Y
13	3PL	317.22	8	77.31	379.48	Y
14	3PL	549.56	8	135.39	379.50	Y
15	3PL	366.51	8	89.63	379.67	Y
16	3PL	567.13	8	139.78	379.71	Y
18	3PL	127.88	8	29.97	379.57	Y
19	3PL	167.87	8	39.97	379.61	Y
20	3PL	96.33	8	22.08	379.56	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
21	3PL	255.05	8	61.76	379.39	Y
22	3PL	865.34	8	214.33	379.83	Y
23	3PL	80.45	8	18.11	379.50	Y
25	3PL	103.14	8	23.79	379.41	Y
26	3PL	358.22	8	87.56	377.55	Y
27	3PL	315.28	8	76.82	379.87	Y
28	3PL	191.30	8	45.82	379.90	Y
29	3PL	137.92	8	32.48	379.76	Y
30	3PL	421.50	8	103.38	379.95	Y
32	3PL	322.08	8	78.52	379.50	Y
33	3PL	715.10	8	176.77	379.63	Y
34	3PL	88.17	8	20.04	379.75	Y
35	3PL	186.75	8	44.69	379.49	Y
36	3PL	196.69	8	47.17	379.61	Y
37	3PL	589.02	8	145.26	380.00	Y
38	3PL	148.90	8	35.22	379.90	Y
39	3PL	88.74	8	20.18	379.52	Y
40	3PL	249.30	8	60.32	379.76	Y
41	3PL	286.00	8	69.50	379.63	Y
42	3PL	83.87	8	18.97	379.85	Y
43	3PL	289.15	8	70.29	379.83	Y
44	3PL	312.41	8	76.10	379.60	Y
45	3PL	95.05	8	21.76	379.66	Y
46	3PL	114.57	8	26.64	379.74	Y
47	3PL	129.90	8	30.48	379.75	Y
50	3PL	711.14	8	175.78	379.52	Y
51	3PL	98.25	8	22.56	378.92	Y
52	2PPC	390.52	17	64.06	377.90	Y
53	2PPC	178.17	17	27.64	379.16	Y
54	2PPC	765.98	17	128.45	377.87	Y
55	2PPC	169.12	17	26.09	377.86	Y
56	2PPC	159.59	17	24.45	377.69	Y
57	2PPC	204.28	17	32.12	375.11	Y
58	2PPC	211.43	26	25.71	377.55	Y
59	2PPC	116.75	26	12.58	376.57	Y
60	2PPC	61.04	26	4.86	374.18	Y
61	2PPC	84.50	26	8.11	375.33	Y

Table O12. Mathematics Grade 8 Item Fit Statistics

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
1	3PL	248.33	8	60.08	286.54	Y
2	3PL	157.31	8	37.33	286.42	Y
3	3PL	170.30	8	40.57	286.17	Y
4	3PL	65.41	8	14.35	286.30	Y
5	3PL	83.49	8	18.87	286.40	Y
6	3PL	978.04	8	242.51	286.31	Y
7	3PL	881.22	8	218.31	286.16	Y
8	3PL	100.62	8	23.15	286.10	Y
9	3PL	84.43	8	19.11	286.17	Y
10	3PL	57.30	8	12.32	286.17	Y
12	3PL	90.97	8	20.74	286.38	Y
14	3PL	494.77	8	121.69	286.09	Y
15	3PL	101.46	8	23.37	286.38	Y
17	3PL	74.64	8	16.66	286.34	Y
18	3PL	70.32	8	15.58	285.99	Y
19	3PL	724.47	8	179.12	286.30	Y
20	3PL	66.60	8	14.65	286.21	Y
22	3PL	97.91	8	22.48	286.15	Y
23	3PL	162.34	8	38.58	286.09	Y
24	3PL	95.82	8	21.96	285.94	Y
25	3PL	365.17	8	89.29	286.12	Y
26	3PL	431.52	8	105.88	285.81	Y
27	3PL	1323.8	8	328.95	286.26	N
28	3PL	81.02	8	18.25	286.16	Y
29	3PL	997.50	8	247.37	286.49	Y
30	3PL	78.38	8	17.59	277.12	Y
31	3PL	93.69	8	21.42	286.37	Y
32	3PL	234.48	8	56.62	286.05	Y
34	3PL	526.19	8	129.55	286.46	Y
35	3PL	106.60	8	24.65	286.14	Y
36	3PL	721.36	8	178.34	286.29	Y
37	3PL	327.34	8	79.83	286.31	Y
38	3PL	71.71	8	15.93	286.39	Y
39	3PL	137.39	8	32.35	286.09	Y
40	3PL	181.04	8	43.26	286.25	Y
42	3PL	905.19	8	224.30	286.17	Y
43	3PL	179.35	8	42.84	286.30	Y

Item	Model	Chi Square	DF	Z-observed	Z-critical	Fit OK?
44	3PL	127.20	8	29.80	286.00	Y
45	3PL	100.58	8	23.15	286.33	Y
46	3PL	564.05	8	139.01	286.04	Y
47	3PL	116.32	8	27.08	286.36	Y
48	3PL	151.40	8	35.85	286.27	Y
50	3PL	107.25	8	24.81	286.21	Y
51	3PL	109.80	8	25.45	286.15	Y
52	2PPC	147.49	17	22.38	282.96	Y
53	2PPC	212.27	17	33.49	284.10	Y
54	2PPC	34.74	17	3.04	281.55	Y
55	2PPC	58.54	17	7.12	279.97	Y
56	2PPC	52.26	17	6.05	279.01	Y
57	2PPC	23.86	17	1.18	280.89	Y
58	2PPC	78.92	26	7.34	280.60	Y
59	2PPC	211.00	26	25.65	282.13	Y
60	2PPC	35.97	26	1.38	274.12	Y
61	2PPC	300.10	26	38.01	279.35	Y

Table O13. ELA Grade 3 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.751	-1.392	0.062		
2	1	0.865	0.524	0.279		
3	1	0.722	-0.039	0.203		
4	1	0.594	0.013	0.064		
5	1	0.784	-0.131	0.129		
6	1	0.397	-1.239	0.020		
7	1	0.714	0.684	0.202		
8	1	0.624	-0.033	0.122		
9	1	0.653	0.168	0.077		
10	1	0.426	0.774	0.081		
11	1	0.919	-0.441	0.110		
12	1	0.661	-0.508	0.024		
13	1	0.635	0.430	0.131		
14	1	0.914	-0.089	0.230		
15	1	0.481	1.259	0.131		
16	1	1.071	0.475	0.218		
17	1	0.427	0.092	0.072		
18	1	1.190	0.486	0.141		

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
19	1	0.982	0.075	0.234		
20	1	0.711	-0.291	0.148		
21	1	0.416	-0.085	0.024		
22	1	0.781	0.850	0.195		
23	1	0.905	0.258	0.155		
24	1	0.709	0.585	0.196		
25	1	0.627	0.073	0.106		
26	2	1.582	-2.318	1.661		
27	2	1.848	-2.531	1.208		
28	4	1.670	-1.930	0.440	2.337	3.989
29	2	1.543	-1.962	2.008		
30	2	1.729	-0.771	2.527		
31	2	1.717	-0.732	2.720		
32	2	1.789	-1.467	2.968		
33	2	1.623	-0.762	2.239		
34	4	1.728	-1.511	0.600	2.788	4.514

Table O14. ELA Grade 4 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.620	-1.192	0.119		
2	1	0.770	0.830	0.353		
3	1	0.911	-0.376	0.267		
4	1	1.265	-0.091	0.262		
5	1	1.007	-0.278	0.154		
6	1	0.505	1.982	0.227		
7	1	0.986	0.588	0.180		
8	1	1.189	0.428	0.223		
9	1	0.850	-0.237	0.199		
10	1	0.721	0.039	0.201		
11	1	0.984	0.250	0.198		
12	1	0.820	-0.932	0.174		
13	1	0.938	-0.081	0.224		
14	1	0.775	0.964	0.164		
15	1	0.519	2.374	0.202		
16	1	0.701	1.060	0.267		
17	1	0.682	1.429	0.207		
18	1	0.366	-0.814	0.015		
19	1	0.475	0.295	0.037		

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
20	1	0.754	0.453	0.204		
21	1	0.395	2.164	0.163		
22	1	0.910	0.838	0.177		
23	1	0.672	0.431	0.243		
24	1	0.806	0.634	0.222		
25	1	0.510	-0.564	0.104		
26	2	1.172	-2.113	0.555		
27	2	1.314	-1.911	1.366		
28	4	1.436	-1.471	-0.219	1.613	3.420
29	2	1.139	-1.842	0.945		
30	2	1.071	-0.455	1.555		
31	2	1.677	-1.723	1.191		
32	2	1.306	-1.102	0.945		
33	2	1.229	-1.882	1.409		
34	4	1.483	-1.548	-0.114	1.620	3.024

Table O15. ELA Grade 5 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.300	-0.677	0.036		
2	1	0.544	0.244	0.139		
3	1	1.025	-1.476	0.249		
4	1	0.338	0.675	0.027		
5	1	0.851	-1.391	0.172		
6	1	0.359	-0.505	0.024		
7	1	0.609	-0.942	0.107		
8	1	0.520	0.347	0.180		
9	1	0.678	-0.006	0.193		
10	1	0.687	0.603	0.325		
11	1	0.670	-1.278	0.077		
12	1	0.788	1.352	0.252		
13	1	0.951	-0.821	0.239		
14	1	0.322	-1.826	0.030		
15	1	0.541	0.342	0.191		
16	1	0.871	-0.088	0.265		
17	1	0.480	1.276	0.305		
18	1	0.467	0.319	0.141		
19	1	0.737	0.090	0.207		
20	1	0.959	-0.020	0.193		

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
21	1	0.742	0.114	0.167		
22	1	0.486	0.784	0.167		
23	1	0.364	0.156	0.119		
24	1	0.615	1.268	0.218		
25	1	0.533	0.660	0.169		
26	1	0.720	0.928	0.257		
27	1	0.567	0.035	0.127		
28	1	0.675	1.192	0.226		
29	1	0.411	1.452	0.426		
30	1	0.845	-0.343	0.175		
31	1	0.736	1.382	0.284		
32	1	0.326	1.902	0.213		
33	1	0.378	-1.608	0.046		
34	1	0.183	-0.261	0.028		
35	1	0.402	-0.494	0.038		
36	2	1.036	-2.278	0.348		
37	2	1.123	-2.443	0.620		
38	4	1.031	-1.361	0.099	1.533	2.920
39	2	1.210	-1.350	0.612		
40	2	1.433	-1.864	0.599		
41	2	1.401	-2.287	0.571		
42	2	1.076	-1.929	0.774		
43	2	1.443	-1.992	1.068		
44	4	1.366	-1.898	-0.294	1.307	3.067

Table O16. ELA Grade 6 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.682	-1.451	0.111		
2	1	0.733	0.250	0.346		
3	1	0.731	0.172	0.244		
4	1	0.793	0.064	0.328		
5	1	0.977	0.375	0.216		
6	1	0.880	0.218	0.196		
7	1	0.636	-1.313	0.039		
8	1	0.777	-1.331	0.225		
9	1	0.854	0.178	0.139		
10	1	0.827	0.654	0.247		
11	1	0.384	0.179	0.061		

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
12	1	0.749	0.051	0.141		
13	1	0.341	1.018	0.098		
14	1	0.538	-0.945	0.056		
15	1	0.531	1.118	0.231		
16	1	0.598	0.958	0.166		
17	1	0.892	-0.160	0.227		
18	1	0.682	0.254	0.179		
19	1	0.922	0.187	0.156		
20	1	0.613	-0.981	0.038		
21	1	0.552	-0.299	0.077		
22	1	1.141	-0.307	0.284		
23	1	1.050	0.046	0.187		
24	1	1.066	0.559	0.210		
25	1	0.966	0.348	0.272		
26	1	0.767	-0.068	0.230		
27	1	0.747	0.320	0.170		
28	1	0.298	0.012	0.023		
29	1	0.711	-0.406	0.158		
30	1	0.451	-1.235	0.051		
31	1	1.018	-0.445	0.171		
32	1	0.614	-0.937	0.118		
33	1	0.951	-0.456	0.177		
34	1	0.561	-1.255	0.022		
35	1	0.737	-0.617	0.215		
36	2	1.639	-3.226	-0.045		
37	2	1.422	-2.880	0.153		
38	4	1.438	-3.446	-1.739	0.217	2.050
39	2	1.556	-2.793	-0.177		
40	2	1.560	-2.357	0.149		
41	2	1.635	-3.037	-0.153		
42	2	1.502	-2.880	0.033		
43	2	1.732	-2.964	-0.106		
44	4	1.409	-2.519	-0.749	0.889	2.235

Table O17. ELA Grade 7 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	1.004	-1.134	0.271		
2	1	0.715	0.138	0.327		
3	1	0.191	-0.493	0.031		
4	1	0.421	-1.062	0.023		
5	1	0.744	0.052	0.408		
6	1	0.738	0.215	0.223		
7	1	0.931	0.798	0.255		
8	1	0.746	0.717	0.263		
9	1	0.330	-1.550	0.035		
10	1	0.802	-1.563	0.095		
11	1	0.909	0.491	0.285		
12	1	1.462	0.380	0.205		
13	1	0.713	0.268	0.127		
14	1	0.377	-0.693	0.028		
15	1	1.154	-0.009	0.248		
16	1	1.511	-0.699	0.297		
17	1	1.330	0.007	0.281		
18	1	0.669	0.549	0.137		
19	1	1.321	0.866	0.197		
20	1	1.006	-0.008	0.233		
21	1	0.944	0.362	0.193		
22	1	0.809	1.178	0.203		
23	1	1.303	0.329	0.242		
24	1	1.171	1.234	0.242		
25	1	1.185	1.204	0.243		
26	1	0.647	0.655	0.204		
27	1	1.006	0.750	0.201		
28	1	1.151	0.288	0.266		
29	1	1.058	1.034	0.183		
30	1	0.864	0.974	0.298		
31	1	0.739	0.115	0.204		
32	1	0.775	0.822	0.235		
33	1	0.640	0.079	0.173		
34	1	0.651	-0.220	0.208		
35	1	0.520	1.725	0.280		
36	2	1.607	-3.105	-0.380		

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
37	2	1.750	-3.204	-0.324		
38	4	1.520	-2.337	-0.889	0.981	2.421
39	2	1.462	-1.761	0.601		
40	2	2.017	-2.486	0.291		
41	2	2.023	-4.159	-0.399		
42	2	1.565	-1.997	-0.055		
43	2	1.858	-2.478	0.469		
44	4	1.529	-3.073	-1.502	0.353	1.801

Table O18. ELA Grade 8 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
1	1	0.612	0.036	0.382		
2	1	0.200	-1.459	0.050		
3	1	1.102	-0.575	0.238		
4	1	0.997	-0.070	0.204		
5	1	1.160	-0.506	0.332		
6	1	0.584	-1.827	0.036		
7	1	1.096	-0.561	0.204		
8	1	0.964	-0.573	0.181		
9	1	0.230	-0.927	0.026		
10	1	0.778	-0.241	0.222		
11	1	0.918	0.333	0.191		
12	1	0.894	-0.537	0.150		
13	1	0.587	-0.520	0.145		
14	1	0.539	-0.513	0.103		
15	1	0.784	0.727	0.183		
16	1	0.798	-0.594	0.226		
17	1	0.572	1.208	0.222		
18	1	1.315	0.379	0.193		
19	1	1.522	0.881	0.232		
20	1	1.119	1.041	0.229		
21	1	1.265	0.660	0.245		
22	1	0.904	-0.283	0.234		
23	1	0.701	0.908	0.277		
24	1	0.595	0.059	0.218		
25	1	1.012	-0.388	0.203		
26	1	1.243	0.103	0.211		
27	1	1.395	-0.031	0.262		

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3	step4
28	1	1.137	0.849	0.234		
29	1	0.687	0.278	0.229		
30	1	0.324	-1.833	0.022		
31	1	0.646	1.842	0.362		
32	1	0.810	-0.214	0.160		
33	1	1.105	-0.435	0.238		
34	1	0.143	0.276	0.070		
35	1	0.872	-0.366	0.144		
36	2	1.800	-4.783	-2.193		
37	2	1.617	-2.930	-0.779		
38	4	1.427	-3.025	-1.836	0.116	1.883
39	2	1.431	-3.201	-0.340		
40	2	1.142	-1.493	0.313		
41	2	1.482	-2.709	-0.060		
42	2	1.693	-3.486	-0.419		
43	2	1.493	-2.430	0.338		
44	4	1.499	-2.970	-1.905	-0.069	1.801

Table O19. Mathematics Grade 3 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.912	-0.552	0.216	
2	1	0.955	-1.635	0.115	
3	1	0.428	-2.399	0.037	
4	1	1.227	0.964	0.275	
5	1	0.903	0.270	0.117	
6	1	0.986	0.052	0.137	
7	1	0.885	-0.188	0.153	
8	1	0.677	-0.492	0.222	
9	1	1.063	-0.949	0.089	
10	1	1.047	0.074	0.317	
11	1	1.359	-0.010	0.177	
12	1	0.969	0.785	0.120	
13	1	0.841	-1.703	0.023	
14	1	1.065	0.889	0.143	
15	1	0.496	-1.370	0.052	
16	1	0.960	1.319	0.174	
17	1	0.418	0.736	0.011	
18	1	1.358	0.395	0.070	

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
19	1	0.668	-2.414	0.270	
20	1	0.781	-0.009	0.372	
21	1	0.795	-1.514	0.033	
22	1	0.871	-0.052	0.118	
23	1	0.891	-1.859	0.099	
24	1	0.965	0.079	0.081	
25	1	0.905	0.188	0.164	
26	1	1.078	0.224	0.154	
27	1	1.237	-0.481	0.197	
28	1	1.254	0.605	0.130	
29	1	1.046	-0.032	0.196	
30	1	0.675	-0.955	0.133	
31	1	1.100	0.193	0.144	
32	1	1.038	0.583	0.303	
33	1	0.697	-0.151	0.122	
34	1	0.936	-0.586	0.115	
35	1	0.777	-0.739	0.144	
36	1	0.762	-1.535	0.205	
37	1	1.236	0.444	0.130	
38	2	0.862	0.879	-0.323	
39	2	0.969	0.914	-1.502	
40	2	0.737	0.213	-0.363	
41	2	0.697	0.487	0.017	
42	2	1.425	-0.440	0.145	
43	3	0.765	-0.968	0.124	0.184
44	3	0.802	1.025	1.463	0.247
45	3	1.093	0.437	0.696	0.049

Table O20. Mathematics Grade 4 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.737	-0.620	0.174	
2	1	0.708	-0.832	0.251	
3	1	0.849	-0.618	0.200	
4	1	0.999	0.675	0.348	
5	1	0.986	-0.575	0.179	
6	1	1.284	0.031	0.112	
7	1	1.109	0.061	0.185	
8	1	0.451	-0.407	0.037	

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
9	1	1.040	-0.453	0.189	
10	1	1.554	1.222	0.132	
11	1	1.172	0.608	0.088	
12	1	1.043	-0.096	0.340	
13	1	0.683	0.308	0.151	
14	1	1.142	-0.116	0.311	
15	1	0.776	-0.326	0.131	
16	1	0.629	-1.269	0.068	
17	1	0.841	0.477	0.204	
18	1	0.946	-0.535	0.130	
19	1	0.931	-1.352	0.017	
20	1	0.894	-0.660	0.375	
21	1	0.824	-1.869	0.040	
22	1	0.827	-2.375	0.026	
23	1	0.837	1.320	0.146	
24	1	0.866	-0.565	0.257	
25	1	0.814	-0.002	0.258	
26	1	1.247	-0.664	0.170	
27	1	0.825	-2.241	0.027	
28	1	0.739	-0.669	0.072	
29	1	0.872	0.934	0.220	
30	1	1.263	0.244	0.204	
31	1	0.853	-0.109	0.315	
32	1	0.719	-1.767	0.204	
33	1	0.743	0.193	0.167	
34	1	0.912	0.214	0.325	
35	1	1.097	0.057	0.157	
36	1	0.725	-0.618	0.198	
37	1	1.148	-0.658	0.172	
38	1	1.156	0.880	0.133	
39	2	0.790	0.155	0.716	
40	2	1.236	-0.074	0.298	
41	2	0.906	0.215	-0.338	
42	2	1.298	-0.045	0.702	
43	2	1.435	1.125	1.416	
44	2	1.031	-1.301	-1.189	
45	3	0.952	-0.585	0.181	0.491
46	3	1.085	0.358	0.338	0.436

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
47	3	0.882	0.410	-0.142	0.038
48	3	1.015	-1.204	0.521	-0.073

Table O21. Mathematics Grade 5 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.818	-1.425	0.039	
2	1	1.315	-0.763	0.196	
3	1	1.047	-0.604	0.129	
4	1	0.983	-0.562	0.155	
5	1	1.081	-0.085	0.289	
6	1	0.738	-0.469	0.146	
7	1	1.107	1.361	0.222	
8	1	0.995	-0.629	0.051	
9	1	1.707	0.795	0.311	
10	1	0.738	1.547	0.219	
11	1	0.687	0.022	0.225	
12	1	1.608	0.886	0.244	
13	1	1.060	-0.422	0.217	
14	1	0.993	-0.022	0.299	
15	1	0.837	0.322	0.340	
16	1	0.975	-0.588	0.193	
17	1	1.624	0.548	0.354	
18	1	1.020	0.357	0.059	
19	1	0.574	-2.162	0.062	
20	1	1.323	0.761	0.157	
21	1	1.102	0.315	0.191	
22	1	0.551	-0.057	0.272	
23	1	1.266	0.508	0.305	
24	1	1.578	0.589	0.195	
25	1	1.070	0.611	0.303	
26	1	0.858	-0.223	0.192	
27	1	1.411	-0.300	0.295	
28	1	1.147	-0.706	0.172	
29	1	0.869	-0.611	0.085	
30	1	1.314	-0.273	0.212	
31	1	1.122	0.545	0.261	
32	1	1.133	0.320	0.163	
33	1	0.729	1.199	0.190	

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
34	1	1.588	0.176	0.150	
35	1	1.012	1.740	0.339	
36	1	0.832	0.319	0.209	
37	1	1.105	0.339	0.300	
38	1	1.006	-0.657	0.109	
39	2	1.273	1.344	1.595	
40	2	1.757	-0.034	1.054	
41	2	1.472	1.232	2.202	
42	2	1.143	0.502	0.006	
43	2	1.543	0.159	0.828	
44	2	1.095	-0.560	-0.298	
45	3	1.208	2.279	1.609	1.395
46	3	1.290	0.663	1.027	0.882
47	3	1.238	0.546	0.716	1.742
48	3	0.961	0.743	0.998	-0.068

Table O22. Mathematics Grade 6 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.971	0.091	0.084	
2	1	0.947	0.336	0.215	
3	1	0.762	1.785	0.209	
4	1	0.948	-0.270	0.224	
5	1	0.616	-1.302	0.172	
6	1	1.351	0.727	0.161	
7	1	0.788	2.453	0.154	
8	1	0.646	1.262	0.134	
9	1	0.699	-0.345	0.309	
10	1	0.970	-0.481	0.230	
11	1	1.249	0.066	0.257	
12	1	1.217	-0.028	0.230	
13	1	0.907	1.108	0.241	
14	1	1.285	1.876	0.110	
15	1	0.570	1.031	0.291	
16	1	1.263	1.195	0.234	
17	1	1.087	1.418	0.273	
18	1	0.948	-0.205	0.166	
19	1	0.940	-0.056	0.202	
20	1	1.326	0.709	0.215	

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
21	1	1.111	1.052	0.182	
22	1	0.804	0.552	0.188	
23	1	0.907	-0.783	0.253	
24	1	0.596	-1.893	0.018	
25	1	0.970	0.225	0.090	
26	1	0.525	-1.435	0.017	
27	1	0.842	-0.499	0.225	
28	1	0.883	0.746	0.192	
29	1	0.745	-0.556	0.148	
30	1	0.435	-0.744	0.135	
31	1	0.932	0.468	0.198	
32	1	1.275	0.896	0.258	
33	1	0.666	0.361	0.146	
34	1	0.980	1.893	0.279	
35	1	0.791	-0.941	0.351	
36	1	1.432	0.134	0.164	
37	1	1.254	0.934	0.213	
38	1	0.493	1.040	0.234	
39	1	0.958	1.062	0.225	
40	1	0.925	0.911	0.124	
41	1	0.919	-1.114	0.359	
42	1	0.978	1.081	0.173	
43	1	1.485	1.051	0.119	
44	1	1.236	0.651	0.190	
45	2	1.251	-0.062	1.900	
46	2	0.963	2.523	-1.521	
47	2	1.562	0.995	4.877	
48	2	1.256	-1.123	0.719	
49	2	1.461	0.658	1.688	
50	2	1.352	-0.261	0.276	
51	3	0.839	-0.240	-1.036	-0.549
52	3	0.759	1.777	0.390	-0.756
53	3	1.118	0.530	1.626	2.638
54	3	0.669	-0.315	-0.452	-0.447

Table O23. Mathematics Grade 7 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.702	-0.502	0.298	
2	1	1.394	-0.065	0.142	
3	1	1.178	1.314	0.181	
4	1	1.011	0.838	0.253	
5	1	1.072	1.127	0.171	
6	1	1.305	0.929	0.174	
7	1	1.096	0.586	0.218	
8	1	1.240	0.072	0.162	
9	1	1.044	0.605	0.225	
10	1	1.267	0.485	0.327	
11	1	1.074	0.719	0.371	
12	1	0.816	-0.069	0.252	
13	1	1.047	0.168	0.085	
14	1	1.332	1.696	0.219	
15	1	1.460	0.824	0.257	
16	1	1.765	0.194	0.171	
17	1	1.177	0.672	0.157	
18	1	1.724	1.534	0.166	
19	1	0.712	-0.297	0.052	
20	1	1.310	0.879	0.169	
21	1	1.340	0.238	0.300	
22	1	0.729	-0.196	0.353	
23	1	1.450	-0.016	0.171	
24	1	1.278	0.725	0.208	
25	1	0.662	1.597	0.251	
26	1	1.463	-0.149	0.217	
27	1	0.775	1.568	0.090	
28	1	0.976	0.012	0.244	
29	1	0.967	1.599	0.188	
30	1	1.334	1.589	0.287	
31	1	1.683	1.340	0.188	
32	1	0.798	-0.476	0.302	
33	1	1.717	1.263	0.199	
34	1	1.213	0.765	0.182	
35	1	1.102	0.405	0.211	
36	1	1.492	0.946	0.305	

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
37	1	1.000	0.992	0.214	
38	1	1.032	0.047	0.187	
39	1	2.707	1.226	0.237	
40	1	1.445	0.748	0.306	
41	1	1.525	0.796	0.281	
42	1	1.527	1.010	0.254	
43	1	0.596	-0.059	0.129	
44	1	1.472	0.942	0.190	
45	2	0.997	0.201	0.573	
46	2	1.758	0.101	1.157	
47	2	1.337	1.582	-0.250	
48	2	1.523	1.541	-0.925	
49	2	1.349	0.218	1.744	
50	2	1.618	-0.104	2.066	
51	3	1.022	0.729	1.229	0.958
52	3	1.147	1.403	1.000	0.286
53	3	1.351	1.010	1.822	-0.401
54	3	1.547	1.952	2.152	2.041

Table O24. Mathematics Grade 8 OP Item Parameter Estimates

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
1	1	0.789	0.021	0.316	
2	1	1.187	0.455	0.201	
3	1	0.680	0.928	0.193	
4	1	1.359	0.262	0.319	
5	1	0.850	0.476	0.287	
6	1	0.819	-0.646	0.152	
7	1	2.513	1.378	0.146	
8	1	1.509	0.427	0.229	
9	1	1.555	0.281	0.279	
10	1	1.202	0.503	0.262	
11	1	1.559	0.587	0.276	
12	1	2.285	1.563	0.208	
13	1	1.223	0.383	0.185	
14	1	1.187	0.350	0.249	
15	1	1.052	0.685	0.293	
16	1	0.886	-0.496	0.238	
17	1	0.912	0.939	0.162	

Item	Max Pts	a-par/alpha	b-par/step1	c-par/step2	step3
18	1	1.171	1.222	0.197	
19	1	0.730	0.421	0.167	
20	1	1.134	0.952	0.187	
21	1	0.726	0.080	0.215	
22	1	0.648	-0.211	0.170	
23	1	0.701	-1.431	0.331	
24	1	0.986	0.620	0.213	
25	1	0.652	-0.836	0.203	
26	1	1.075	1.008	0.349	
27	1	1.238	0.704	0.143	
28	1	1.280	0.910	0.188	
29	1	0.574	-0.131	0.187	
30	1	1.179	0.231	0.264	
31	1	0.681	-0.532	0.144	
32	1	1.196	0.144	0.235	
33	1	0.951	0.595	0.243	
34	1	1.365	0.027	0.265	
35	1	0.935	0.149	0.313	
36	1	1.870	1.614	0.174	
37	1	1.852	1.103	0.273	
38	1	1.633	0.905	0.246	
39	1	0.972	0.853	0.300	
40	1	1.291	1.347	0.409	
41	1	1.239	0.084	0.257	
42	1	1.000	0.050	0.352	
43	1	0.968	0.556	0.205	
44	1	1.239	0.263	0.208	
45	2	1.549	0.404	1.175	
46	2	1.297	0.467	-0.026	
47	2	1.269	1.240	0.138	
48	2	1.093	1.145	1.123	
49	2	1.814	0.685	1.282	
50	2	1.666	0.901	0.473	
51	3	1.120	1.459	1.520	1.714
52	3	0.942	1.072	0.957	-0.073
53	3	1.262	2.569	0.933	0.572
54	3	0.803	-1.015	-0.037	0.098

Appendix P: Derivation and Estimation of Classification Consistency and Accuracy

Classification Consistency

Assume that θ is a single latent trait measured by a test and denote Φ as a latent random variable. When a test X consists of K items and its maximum number correct score is N , the marginal probability of the number correct (NC) score x is

$$P(X = x) = \int P(X = x | \Phi = \theta) g(\theta) d\theta, \quad x = 0, 1, \dots, N$$

where

$g(\theta)$ is the density of θ .

In this report, the marginal distribution $P(X = x)$ is denoted as $f(x)$, and the conditional error distribution $P(X = x | \Phi = \theta)$ is denoted as $f(x | \theta)$. It is assumed that examinees are classified into one of H mutually exclusive categories on the basis of predetermined $H - 1$ observed score cutoffs, C_1, C_2, \dots, C_{H-1} . Let L_h represent the h th category into which examinees with $C_{h-1} \leq X < C_h$ are classified. $C_0 = 0$ and $C_H =$ the maximum number-correct score plus one. Then, the conditional and marginal probabilities of each category classification are as follows:

$$P(X \in L_h | \theta) = \sum_{x=C_{h-1}}^{C_h-1} f(x | \theta), \quad h = 1, 2, \dots, H$$

$$P(X \in L_h) = \int \sum_{x=C_{h-1}}^{C_h-1} f(x | \theta) g(\theta) d\theta, \quad h = 1, 2, \dots, H$$

Because obtaining test scores from two independent administrations of New York State tests was not feasible due to item release after each OP administration, a psychometric model was used to obtain the estimated classification consistency indices using test scores from a single administration. Based on the psychometric model, a symmetric H -by- H contingency table can be constructed. The elements of the H -by- H contingency table consist of the joint probabilities of the row and column observed category classifications.

That two administrations are independent implies that if X_1 and X_2 represent the raw score random variables on the two administrations, then, conditioned on θ , X_1 and X_2 are independent and identically distributed. Consequently, the conditional bivariate distribution of X_1 and X_2 is

$$f(x_1, x_2 | \theta) = f(x_1 | \theta) f(x_2 | \theta)$$

The marginal bivariate distribution of X_1 and X_2 can be expressed as follows:

$$f(x_1, x_2) = \int f(x_1, x_2 | \theta) f(\theta) d\theta$$

Consistent classification means that both X_1 and X_2 fall in the same category. The conditional probability of falling in the same category on the two administrations is

$$P(X_1 \in L_h, X_2 \in L_h | \theta) = \left[\sum_{x_1=C_{h-1}}^{C_{h+1}} f(x_1 | \theta) \right]^2, h = 1, 2, \dots, H$$

The agreement index P , conditional on theta, is obtained by

$$P(\theta) = \sum_{h=1}^H P(X_1 \in L_h, X_2 \in L_h | \theta)$$

The agreement index (classification consistency) can be computed as

$$P = \int P(\theta)g(\theta)d(\theta)$$

The probability of consistent classification by chance, P_c , is the sum of squared marginal probabilities of each category classification.

$$P_c = \sum_{h=1}^H P(X_1 \in L_h)P(X_2 \in L_h) = \sum_{h=1}^H [P(X_1 \in L_h)]^2$$

Then, Kappa (Cohen, 1960) is

$$k = \frac{P - P_c}{1 - P_c}$$

Classification Accuracy

Let Γ_w denote true category. When an examinee has an observed score, $x \in L_h$ ($h = 1, 2, \dots, H$), and a latent score, $\theta \in \Gamma_w$ ($w = 1, 2, \dots, H$), an accurate classification is made when $h = w$. The conditional probability of accurate classification is

$$\gamma(\theta) = P(X \in L_w | \theta),$$

where

w is the category such that $\theta \in \Gamma_w$

Lee (2008) thoroughly discusses this IRT method for estimating decision indices, including the computational method used to estimate the results when integrating across the latent variable, θ .

Estimating Classification Indices

The classification consistency and accuracy estimates were obtained using an open-source software program, IRT-CLASS v2.0 (Lee & Kolen, 2006). Below is a brief description of the files that are used and their purpose. (See the IRT-CLASS v2.0 manual for complete instructions.)

Files needed:

- Raw-to-Scale score conversion file
 - a. Contains the raw-to-scale score conversions
 - b. This is used to provide both raw and scale score classification estimates, which is useful when the raw-to-scale score transformation is not one-to-one.
- Cut score file
 - a. Contains the cut scores to be used
 - b. Results are provided for all cut scores simultaneously (all performance levels), as well as the estimates based on each of the cut scores separately (Level III only).
- Item parameter file
 - a. This contains the IRT model used and item parameter estimates.
 - b. This information is used when calculating the classification indices.
- Theta file
 - a. Contains the theta distribution in terms of quadrature points
 - b. The theta and the item parameter files are used to solve the integrals mentioned above.
- Control card
 - a. This is used to run the program.
 - b. Identifies the names of the four files above and gives a name to the output file

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Tables Q1–Q12 show the PBT raw-to-scale score conversion tables, while Tables Q13–Q24 show the CBT raw-to-scale score conversion tables. Tables Q25–Q36 show the scale score distributions that include all students with valid scores, by frequency (n-count), percent, cumulative frequency, and cumulative percent.

Table Q1. PBT ELA Grade 3 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	180	52	24	306	8
1	188	43	25	309	8
2	196	36	26	312	8
3	204	30	27	315	8
4	212	25	28	320	8
5	220	21	29	321	8
6	228	18	30	325	8
7	237	15	31	328	8
8	245	13	32	331	9
9	251	12	33	334	9
10	256	11	34	338	9
11	261	11	35	341	9
12	266	10	36	345	9
13	270	10	37	349	9
14	273	10	38	353	9
15	277	10	39	358	10
16	281	9	40	362	10
17	284	9	41	367	11
18	287	9	42	373	12
19	291	9	43	379	12
20	294	9	44	388	14
21	297	9	45	396	16
22	300	9	46	404	19
23	303	8	47	412	22

Table Q2. PBT ELA Grade 4 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	164	53	24	302	9
1	172	46	25	305	9

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
2	180	40	26	308	9
3	188	35	27	311	9
4	196	30	28	314	9
5	204	26	29	320	9
6	212	23	30	321	9
7	220	21	31	324	9
8	228	18	32	327	9
9	237	16	33	331	9
10	244	15	34	334	10
11	251	13	35	338	10
12	256	13	36	343	10
13	261	12	37	346	10
14	266	12	38	351	11
15	270	11	39	356	11
16	274	11	40	361	12
17	278	10	41	367	13
18	282	10	42	374	14
19	287	10	43	382	15
20	289	10	44	392	18
21	292	9	45	400	20
22	295	9	46	408	23
23	299	9	47	416	26

Table Q3. PBT ELA Grade 5 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	126	67	29	291	10
1	134	59	30	295	10
2	142	53	31	298	10
3	150	47	32	301	10
4	158	42	33	304	10
5	166	37	34	307	10
6	174	33	35	310	10
7	182	29	36	314	10
8	190	26	37	317	10
9	198	23	38	320	10
10	206	20	39	323	10

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
11	214	18	40	327	10
12	222	16	41	330	10
13	229	15	42	334	11
14	234	14	43	338	11
15	240	13	44	342	11
16	245	13	45	346	11
17	249	12	46	350	12
18	253	12	47	355	12
19	257	11	48	360	13
20	261	11	49	365	13
21	265	11	50	371	14
22	268	11	51	378	15
23	272	11	52	386	17
24	275	10	53	396	19
25	279	10	54	404	21
26	282	10	55	412	23
27	285	10	56	420	26
28	289	10	57	428	28

Table Q4. PBT ELA Grade 6 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	128	81	29	283	8
1	136	68	30	285	8
2	144	57	31	288	8
3	152	48	32	290	8
4	160	40	33	293	8
5	168	34	34	295	8
6	176	29	35	298	8
7	184	24	36	300	8
8	192	21	37	303	8
9	201	18	38	305	8
10	209	15	39	308	8
11	217	14	40	310	8
12	225	12	41	313	8
13	230	12	42	316	8
14	234	11	43	320	9

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
15	239	11	44	322	9
16	243	10	45	325	9
17	246	10	46	328	9
18	250	10	47	332	9
19	253	10	48	338	10
20	257	10	49	340	10
21	260	9	50	344	11
22	263	9	51	350	12
23	266	9	52	356	13
24	269	9	53	363	14
25	272	9	54	373	17
26	275	9	55	387	22
27	277	9	56	395	25
28	280	8	57	403	29

Table Q5. PBT ELA Grade 7 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	133	89	29	291	8
1	140	78	30	293	8
2	148	67	31	296	8
3	156	57	32	298	8
4	164	48	33	301	8
5	172	41	34	303	8
6	180	34	35	305	7
7	188	28	36	308	7
8	196	24	37	310	7
9	204	20	38	313	7
10	212	17	39	315	7
11	220	15	40	318	8
12	228	13	41	320	8
13	235	12	42	323	8
14	240	11	43	326	8
15	245	11	44	328	8
16	249	10	45	331	8
17	253	10	46	334	8
18	257	10	47	338	9

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
19	261	9	48	341	9
20	264	9	49	347	9
21	268	9	50	349	10
22	271	9	51	354	10
23	274	9	52	359	11
24	277	8	53	366	13
25	280	8	54	374	15
26	283	8	55	386	19
27	287	8	56	394	22
28	288	8	57	402	26

Table Q6. PBT ELA Grade 8 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	121	82	29	281	8
1	129	71	30	284	8
2	137	62	31	286	8
3	145	53	32	288	8
4	153	45	33	291	8
5	161	38	34	293	8
6	168	33	35	296	8
7	176	28	36	298	8
8	184	24	37	301	8
9	192	20	38	304	8
10	200	18	39	306	8
11	208	15	40	309	8
12	216	14	41	311	8
13	224	12	42	316	8
14	229	12	43	317	8
15	234	11	44	320	8
16	239	11	45	323	8
17	243	10	46	326	9
18	247	10	47	330	9
19	251	10	48	334	9
20	254	10	49	338	10
21	257	9	50	343	10
22	261	9	51	347	11

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
23	264	9	52	354	12
24	267	9	53	361	14
25	270	9	54	371	17
26	273	9	55	386	23
27	275	8	56	394	26
28	278	8	57	402	30

Table Q7. PBT Mathematics Grade 3 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	145	74	29	295	8
1	153	65	30	297	8
2	161	56	31	299	8
3	169	49	32	302	7
4	177	42	33	304	7
5	185	36	34	306	7
6	193	31	35	308	7
7	201	27	36	310	7
8	209	23	37	314	7
9	217	20	38	315	7
10	225	18	39	317	8
11	233	16	40	319	8
12	239	15	41	321	8
13	245	13	42	324	8
14	250	13	43	326	8
15	254	12	44	329	8
16	258	11	45	332	8
17	262	11	46	335	9
18	266	10	47	340	9
19	269	10	48	342	10
20	272	10	49	346	10
21	275	9	50	350	11
22	278	9	51	355	12
23	281	9	52	362	14
24	285	8	53	370	16
25	286	8	54	381	20
26	288	8	55	389	24

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
27	290	8	56	397	28
28	293	8			

Table Q8. PBT Mathematics Grade 4 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	133	91	32	292	8
1	141	76	33	294	7
2	149	64	34	296	7
3	157	54	35	298	7
4	165	45	36	300	7
5	173	39	37	302	7
6	182	32	38	304	7
7	190	28	39	306	7
8	198	24	40	308	7
9	206	21	41	310	7
10	214	19	42	312	7
11	222	17	43	314	7
12	230	15	44	316	7
13	236	14	45	319	7
14	241	13	46	321	8
15	245	13	47	323	8
16	250	12	48	325	8
17	254	11	49	328	8
18	257	11	50	330	8
19	261	10	51	333	8
20	264	10	52	336	9
21	267	9	53	341	9
22	269	9	54	343	9
23	272	9	55	347	10
24	275	9	56	351	11
25	277	8	57	356	11
26	279	8	58	362	13
27	283	8	59	370	15
28	284	8	60	381	19
29	286	8	61	389	22
30	288	8	62	397	26

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
31	290	8			

Table Q9. PBT Mathematics Grade 5 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	151	116	32	307	6
1	159	99	33	309	6
2	167	85	34	311	6
3	175	73	35	313	6
4	183	62	36	314	6
5	191	54	37	316	6
6	199	46	38	319	6
7	207	39	39	320	6
8	215	33	40	321	6
9	223	28	41	323	6
10	232	23	42	325	6
11	242	18	43	327	6
12	250	15	44	329	6
13	256	13	45	331	6
14	261	12	46	332	6
15	265	11	47	334	7
16	269	10	48	337	7
17	273	10	49	339	7
18	276	9	50	341	7
19	279	9	51	343	7
20	282	8	52	346	8
21	284	8	53	349	8
22	287	8	54	352	8
23	289	8	55	355	9
24	291	7	56	359	10
25	294	7	57	363	10
26	296	7	58	369	12
27	298	7	59	375	13
28	300	7	60	385	17
29	302	7	61	393	20
30	304	7	62	401	24
31	305	7			

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Table Q10. PBT Mathematics Grade 6 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	132	105	35	308	7
1	140	92	36	310	7
2	148	80	37	312	7
3	157	69	38	314	7
4	165	61	39	316	7
5	173	54	40	318	7
6	181	47	41	319	7
7	189	42	42	321	7
8	197	37	43	323	7
9	205	32	44	325	7
10	213	28	45	327	7
11	221	25	46	329	7
12	230	21	47	331	7
13	238	19	48	333	7
14	245	16	49	335	7
15	251	15	50	337	7
16	256	14	51	340	7
17	261	13	52	342	7
18	265	12	53	344	8
19	268	11	54	347	8
20	272	11	55	349	8
21	275	10	56	352	8
22	278	10	57	355	8
23	281	10	58	358	9
24	284	9	59	362	9
25	286	9	60	366	10
26	289	9	61	370	10
27	291	8	62	375	11
28	293	8	63	381	12
29	295	8	64	388	14
30	298	8	65	397	16
31	300	8	66	405	18
32	302	8	67	413	20
33	304	8	68	421	23
34	306	7			

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Table Q11. PBT Mathematics Grade 7 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	160	189	35	318	5
1	168	161	36	319	5
2	176	136	37	321	5
3	184	114	38	322	5
4	191	98	39	323	5
5	199	82	40	325	5
6	207	68	41	326	5
7	215	57	42	327	5
8	223	47	43	329	5
9	230	40	44	330	5
10	238	33	45	331	5
11	246	27	46	333	5
12	259	19	47	334	5
13	267	15	48	335	5
14	273	13	49	337	5
15	277	12	50	338	5
16	281	10	51	340	5
17	285	9	52	341	5
18	288	9	53	343	5
19	291	8	54	345	5
20	293	8	55	346	6
21	295	7	56	348	6
22	297	7	57	350	6
23	299	7	58	352	6
24	301	7	59	354	6
25	303	6	60	357	7
26	305	6	61	360	7
27	306	6	62	363	8
28	308	6	63	367	9
29	309	6	64	371	10
30	311	6	65	377	11
31	312	6	66	385	14
32	314	6	67	393	17
33	315	6	68	401	21
34	317	5			

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Table Q12. PBT Mathematics Grade 8 RSSS Table

Raw Score	Scale Score	Standard Error	Raw Score	Scale Score	Standard Error
0	134	173	35	314	6
1	142	154	36	315	6
2	150	137	37	317	6
3	158	121	38	318	6
4	166	107	39	319	6
5	173	95	40	321	6
6	181	83	41	322	6
7	189	72	42	324	6
8	197	62	43	325	6
9	205	53	44	327	6
10	213	45	45	328	6
11	221	38	46	330	6
12	229	32	47	331	6
13	244	24	48	333	6
14	254	19	49	334	6
15	262	16	50	336	6
16	268	14	51	337	6
17	273	13	52	339	6
18	277	12	53	341	6
19	280	11	54	342	6
20	284	10	55	344	6
21	287	9	56	346	6
22	289	9	57	349	6
23	292	8	58	350	6
24	294	8	59	353	7
25	296	8	60	355	7
26	298	7	61	358	7
27	300	7	62	361	8
28	302	7	63	365	9
29	304	7	64	370	10
30	306	6	65	376	12
31	307	6	66	385	15
32	309	6	67	392	19
33	311	6	68	400	23
34	312	6			

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Table Q13. CBT ELA Grade 3 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	184	47	24	310	8
1	192	39	25	313	8
2	200	33	26	316	8
3	208	27	27	319	8
4	216	23	28	324	8
5	224	19	29	325	8
6	232	17	30	329	8
7	241	14	31	332	9
8	249	13	32	335	9
9	255	12	33	338	9
10	260	11	34	342	9
11	265	10	35	345	9
12	270	10	36	349	9
13	274	10	37	353	9
14	277	10	38	357	10
15	281	9	39	362	10
16	285	9	40	366	11
17	288	9	41	371	11
18	291	9	42	377	12
19	295	9	43	383	13
20	298	9	44	392	15
21	301	9	45	400	18
22	304	8	46	408	21
23	307	8	47	412	22

* A CBT mode adjustment has been taken into account for these scale scores

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Table Q14. CBT ELA Grade 4 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	169	49	24	307	9
1	177	42	25	310	9
2	185	37	26	313	9
3	193	32	27	316	9
4	201	28	28	319	9
5	209	24	29	325	9
6	217	21	30	326	9
7	225	19	31	329	9
8	233	17	32	332	9
9	242	15	33	336	10
10	249	14	34	339	10
11	256	13	35	343	10
12	261	12	36	348	11
13	266	12	37	351	11
14	271	11	38	356	11
15	275	11	39	361	12
16	279	10	40	366	13
17	283	10	41	372	14
18	287	10	42	379	15
19	292	9	43	387	16
20	294	9	44	397	19
21	297	9	45	405	22
22	300	9	46	413	24
23	304	9	47	416	26

* A CBT mode adjustment has been taken into account for these scale scores

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Table Q15. CBT ELA Grade 5 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	128	65	29	293	10
1	136	58	30	297	10
2	144	51	31	300	10
3	152	46	32	303	10
4	160	40	33	306	10
5	168	36	34	309	10
6	176	32	35	312	10
7	184	28	36	316	10
8	192	25	37	319	10
9	200	22	38	322	10
10	208	20	39	325	10
11	216	18	40	329	10
12	224	16	41	332	11
13	231	15	42	336	11
14	236	14	43	340	11
15	242	13	44	344	11
16	247	12	45	348	12
17	251	12	46	352	12
18	255	12	47	357	12
19	259	11	48	362	13
20	263	11	49	367	14
21	267	11	50	373	14
22	270	11	51	380	15
23	274	10	52	388	17
24	277	10	53	398	19
25	281	10	54	406	21
26	284	10	55	414	24
27	287	10	56	422	26
28	291	10	57	428	28

* A CBT mode adjustment has been taken into account for these scale scores

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Table Q16. CBT ELA Grade 6 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	133	73	29	288	8
1	141	61	30	290	8
2	149	51	31	293	8
3	157	43	32	295	8
4	165	36	33	298	8
5	173	31	34	300	8
6	181	26	35	303	8
7	189	22	36	305	8
8	197	19	37	308	8
9	206	16	38	310	8
10	214	14	39	313	8
11	222	13	40	315	8
12	230	12	41	318	8
13	235	11	42	321	9
14	239	11	43	325	9
15	244	10	44	327	9
16	248	10	45	330	9
17	251	10	46	333	10
18	255	10	47	337	10
19	258	10	48	343	11
20	262	9	49	345	11
21	265	9	50	349	12
22	268	9	51	355	13
23	271	9	52	361	14
24	274	9	53	368	16
25	277	9	54	378	19
26	280	8	55	392	24
27	282	8	56	400	27
28	285	8	57	403	29

* A CBT mode adjustment has been taken into account for these scale scores

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Table Q17. CBT ELA Grade 7 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	135	85	29	293	8
1	142	75	30	295	8
2	150	64	31	298	8
3	158	55	32	300	8
4	166	46	33	303	8
5	174	39	34	305	7
6	182	33	35	307	7
7	190	27	36	310	7
8	198	23	37	312	7
9	206	19	38	315	7
10	214	16	39	317	7
11	222	14	40	320	8
12	230	13	41	322	8
13	237	11	42	325	8
14	242	11	43	328	8
15	247	10	44	330	8
16	251	10	45	333	8
17	255	10	46	336	8
18	259	9	47	340	9
19	263	9	48	343	9
20	266	9	49	349	10
21	270	9	50	351	10
22	273	9	51	356	11
23	276	9	52	361	12
24	279	8	53	368	13
25	282	8	54	376	16
26	285	8	55	388	20
27	289	8	56	396	23
28	290	8	57	402	26

* A CBT mode adjustment has been taken into account for these scale scores

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Table Q18. CBT ELA Grade 8 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	121	82	29	281	8
1	129	71	30	284	8
2	137	62	31	286	8
3	145	53	32	288	8
4	153	45	33	291	8
5	161	38	34	293	8
6	168	33	35	296	8
7	176	28	36	298	8
8	184	24	37	301	8
9	192	20	38	304	8
10	200	18	39	306	8
11	208	15	40	309	8
12	216	14	41	311	8
13	224	12	42	316	8
14	229	12	43	317	8
15	234	11	44	320	8
16	239	11	45	323	8
17	243	10	46	326	9
18	247	10	47	330	9
19	251	10	48	334	9
20	254	10	49	338	10
21	257	9	50	343	10
22	261	9	51	347	11
23	264	9	52	354	12
24	267	9	53	361	14
25	270	9	54	371	17
26	273	9	55	386	23
27	275	8	56	394	26
28	278	8	57	402	30

* A CBT mode adjustment has been taken into account for these scale scores

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Table Q19. CBT Mathematics Grade 3 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	149	69	29	299	8
1	157	60	30	301	8
2	165	52	31	303	7
3	173	45	32	306	7
4	181	39	33	308	7
5	189	33	34	310	7
6	197	29	35	312	7
7	205	25	36	314	7
8	213	22	37	318	8
9	221	19	38	319	8
10	229	17	39	321	8
11	237	15	40	323	8
12	243	14	41	325	8
13	249	13	42	328	8
14	254	12	43	330	8
15	258	11	44	333	9
16	262	11	45	336	9
17	266	10	46	339	9
18	270	10	47	344	10
19	273	10	48	346	10
20	276	9	49	350	11
21	279	9	50	354	12
22	282	9	51	359	13
23	285	8	52	366	15
24	289	8	53	374	17
25	290	8	54	385	22
26	292	8	55	393	26
27	294	8	56	397	28
28	297	8			

* A CBT mode adjustment has been taken into account for these scale scores

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Table Q20. CBT Mathematics Grade 4 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	138	82	32	297	7
1	146	69	33	299	7
2	154	58	34	301	7
3	162	48	35	303	7
4	170	41	36	305	7
5	178	35	37	307	7
6	187	29	38	309	7
7	195	25	39	311	7
8	203	22	40	313	7
9	211	20	41	315	7
10	219	18	42	317	7
11	227	16	43	319	7
12	235	14	44	321	8
13	241	13	45	324	8
14	246	12	46	326	8
15	250	12	47	328	8
16	255	11	48	330	8
17	259	10	49	333	8
18	262	10	50	335	8
19	266	10	51	338	9
20	269	9	52	341	9
21	272	9	53	346	10
22	274	9	54	348	10
23	277	8	55	352	11
24	280	8	56	356	11
25	282	8	57	361	13
26	284	8	58	367	14
27	288	8	59	375	17
28	289	8	60	386	21
29	291	8	61	394	25
30	293	7	62	397	26
31	295	7			

* A CBT mode adjustment has been taken into account for these scale scores

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Table Q21. CBT Mathematics Grade 5 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	153	112	32	309	6
1	161	95	33	311	6
2	169	82	34	313	6
3	177	70	35	315	6
4	185	60	36	316	6
5	193	52	37	318	6
6	201	44	38	321	6
7	209	38	39	322	6
8	217	32	40	323	6
9	225	27	41	325	6
10	234	22	42	327	6
11	244	18	43	329	6
12	252	15	44	331	6
13	258	13	45	333	7
14	263	11	46	334	7
15	267	11	47	336	7
16	271	10	48	339	7
17	275	9	49	341	7
18	278	9	50	343	7
19	281	8	51	345	8
20	284	8	52	348	8
21	286	8	53	351	8
22	289	8	54	354	9
23	291	7	55	357	9
24	293	7	56	361	10
25	296	7	57	365	11
26	298	7	58	371	12
27	300	7	59	377	14
28	302	7	60	387	17
29	304	7	61	395	21
30	306	7	62	401	24
31	307	6			

* A CBT mode adjustment has been taken into account for these scale scores

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Table Q22. CBT Mathematics Grade 6 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	135	100	35	311	7
1	143	87	36	313	7
2	151	77	37	315	7
3	160	66	38	317	7
4	168	58	39	319	7
5	176	51	40	321	7
6	184	45	41	322	7
7	192	40	42	324	7
8	200	35	43	326	7
9	208	31	44	328	7
10	216	27	45	330	7
11	224	24	46	332	7
12	233	20	47	334	7
13	241	18	48	336	7
14	248	16	49	338	7
15	254	14	50	340	7
16	259	13	51	343	7
17	264	12	52	345	8
18	268	11	53	347	8
19	271	11	54	350	8
20	275	10	55	352	8
21	278	10	56	355	8
22	281	10	57	358	9
23	284	9	58	361	9
24	287	9	59	365	10
25	289	9	60	369	10
26	292	8	61	373	11
27	294	8	62	378	12
28	296	8	63	384	13
29	298	8	64	391	14
30	301	8	65	400	16
31	303	8	66	408	19
32	305	7	67	416	21
33	307	7	68	421	23
34	309	7			

* A CBT mode adjustment has been taken into account for these scale scores

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Table Q23. CBT Mathematics Grade 7 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	160	189	35	318	5
1	168	161	36	319	5
2	176	136	37	321	5
3	184	114	38	322	5
4	191	98	39	323	5
5	199	82	40	325	5
6	207	68	41	326	5
7	215	57	42	327	5
8	223	47	43	329	5
9	230	40	44	330	5
10	238	33	45	331	5
11	246	27	46	333	5
12	259	19	47	334	5
13	267	15	48	335	5
14	273	13	49	337	5
15	277	12	50	338	5
16	281	10	51	340	5
17	285	9	52	341	5
18	288	9	53	343	5
19	291	8	54	345	5
20	293	8	55	346	6
21	295	7	56	348	6
22	297	7	57	350	6
23	299	7	58	352	6
24	301	7	59	354	6
25	303	6	60	357	7
26	305	6	61	360	7
27	306	6	62	363	8
28	308	6	63	367	9
29	309	6	64	371	10
30	311	6	65	377	11
31	312	6	66	385	14
32	314	6	67	393	17
33	315	6	68	401	21
34	317	5			

* A CBT mode adjustment has been taken into account for these scale scores

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Table Q24. CBT Mathematics Grade 8 RSSS Table

Raw Score	Scale Score*	Standard Error	Raw Score	Scale Score*	Standard Error
0	142	154	35	322	6
1	150	137	36	323	6
2	158	121	37	325	6
3	166	107	38	326	6
4	174	94	39	327	6
5	181	83	40	329	6
6	189	72	41	330	6
7	197	62	42	332	6
8	205	53	43	333	6
9	213	45	44	335	6
10	221	38	45	336	6
11	229	32	46	338	6
12	237	28	47	339	6
13	252	20	48	341	6
14	262	16	49	342	6
15	270	14	50	344	6
16	276	12	51	345	6
17	281	11	52	347	6
18	285	10	53	349	6
19	288	9	54	350	6
20	292	8	55	352	6
21	295	8	56	354	7
22	297	7	57	357	7
23	300	7	58	358	7
24	302	7	59	361	8
25	304	7	60	363	8
26	306	6	61	366	9
27	308	6	62	369	10
28	310	6	63	373	11
29	312	6	64	378	13
30	314	6	65	384	15
31	315	6	66	393	19
32	317	6	67	400	23
33	319	6	68	400	23
34	320	6			

* A CBT mode adjustment has been taken into account for these scale scores

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Table Q25. PBT ELA Grade 3 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
180	24	0.01%	24	0.01%
188	73	0.04%	97	0.05%
192	2	0.00%	99	0.05%
196	173	0.10%	272	0.15%
200	1	0.00%	273	0.15%
204	427	0.23%	700	0.38%
208	7	0.00%	707	0.39%
212	825	0.45%	1,532	0.84%
216	12	0.01%	1,544	0.85%
220	1,321	0.73%	2,865	1.58%
224	22	0.01%	2,887	1.59%
228	1,808	0.99%	4,695	2.58%
232	43	0.02%	4,738	2.61%
237	2,312	1.27%	7,050	3.88%
241	60	0.03%	7,110	3.91%
245	2,790	1.53%	9,900	5.44%
249	65	0.04%	9,965	5.48%
251	2,980	1.64%	12,945	7.12%
255	83	0.05%	13,028	7.16%
256	3,248	1.79%	16,276	8.95%
260	95	0.05%	16,371	9.00%
261	3,650	2.01%	20,021	11.0%
265	79	0.04%	20,100	11.1%
266	3,833	2.11%	23,933	13.2%
270	4,084	2.25%	28,017	15.4%
273	4,128	2.27%	32,145	17.7%
274	108	0.06%	32,253	17.7%
277	4,313	2.37%	36,566	20.1%
281	4,518	2.48%	41,084	22.6%
284	4,691	2.58%	45,775	25.2%
285	117	0.06%	45,892	25.2%
287	4,782	2.63%	50,674	27.9%
288	121	0.07%	50,795	27.9%
291	5,165	2.84%	55,960	30.8%
294	5,257	2.89%	61,217	33.7%
295	137	0.08%	61,354	33.7%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
297	5,550	3.05%	66,904	36.8%
298	153	0.08%	67,057	36.9%
300	5,539	3.05%	72,596	39.9%
301	115	0.06%	72,711	40.0%
303	5,716	3.14%	78,427	43.1%
304	156	0.09%	78,583	43.2%
306	5,916	3.25%	84,499	46.5%
307	155	0.09%	84,654	46.6%
309	5,950	3.27%	90,604	49.8%
310	159	0.09%	90,763	49.9%
312	6,267	3.45%	97,030	53.4%
313	155	0.09%	97,185	53.4%
315	6,316	3.47%	103,501	56.9%
316	150	0.08%	103,651	57.0%
319	175	0.10%	103,826	57.1%
320	6,447	3.55%	110,273	60.6%
321	6,719	3.69%	116,992	64.3%
324	167	0.09%	117,159	64.4%
325	6,779	3.73%	123,938	68.2%
328	6,479	3.56%	130,417	71.7%
329	144	0.08%	130,561	71.8%
331	6,552	3.60%	137,113	75.4%
332	129	0.07%	137,242	75.5%
334	6,062	3.33%	143,304	78.8%
335	147	0.08%	143,451	78.9%
338	6,070	3.34%	149,521	82.2%
341	5,538	3.05%	155,059	85.3%
342	102	0.06%	155,161	85.3%
345	5,136	2.82%	160,297	88.2%
349	4,664	2.56%	164,961	90.7%
353	4,035	2.22%	168,996	92.9%
357	57	0.03%	169,053	93.0%
358	3,482	1.91%	172,535	94.9%
362	2,837	1.56%	175,372	96.4%
366	23	0.01%	175,395	96.5%
367	2,217	1.22%	177,612	97.7%
371	15	0.01%	177,627	97.7%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
373	1,677	0.92%	179,304	98.6%
377	10	0.01%	179,314	98.6%
379	1,163	0.64%	180,477	99.2%
383	6	0.00%	180,483	99.3%
388	766	0.42%	181,249	99.7%
392	1	0.00%	181,250	99.7%
396	385	0.21%	181,635	99.9%
404	157	0.09%	181,792	100%
408	1	0.00%	181,793	100%
412	48	0.03%	181,841	100%

Table Q26. PBT ELA Grade 4 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
164	6	0.00%	6	0.00%
172	27	0.01%	33	0.02%
177	1	0.00%	34	0.02%
180	75	0.04%	109	0.06%
188	195	0.11%	304	0.17%
193	2	0.00%	306	0.17%
196	359	0.20%	665	0.37%
201	3	0.00%	668	0.37%
204	652	0.36%	1,320	0.73%
209	13	0.01%	1,333	0.73%
212	1,028	0.57%	2,361	1.30%
217	23	0.01%	2,384	1.31%
220	1,377	0.76%	3,761	2.07%
225	21	0.01%	3,782	2.08%
228	1,760	0.97%	5,542	3.05%
233	34	0.02%	5,576	3.07%
237	2,192	1.21%	7,768	4.27%
242	43	0.02%	7,811	4.30%
244	2,559	1.41%	10,370	5.70%
249	47	0.03%	10,417	5.73%
251	2,952	1.62%	13,369	7.35%
256	3,327	1.83%	16,696	9.18%
261	3,863	2.13%	20,559	11.3%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
266	4,108	2.26%	24,667	13.6%
270	4,229	2.33%	28,896	15.9%
271	79	0.04%	28,975	15.9%
274	4,554	2.51%	33,529	18.4%
275	89	0.05%	33,618	18.5%
278	4,898	2.69%	38,516	21.2%
279	94	0.05%	38,610	21.2%
282	5,179	2.85%	43,789	24.1%
283	90	0.05%	43,879	24.1%
287	5,350	2.94%	49,229	27.1%
289	5,563	3.06%	54,792	30.1%
292	5,848	3.22%	60,640	33.4%
294	98	0.05%	60,738	33.4%
295	6,005	3.30%	66,743	36.7%
297	105	0.06%	66,848	36.8%
299	6,152	3.38%	73,000	40.2%
300	105	0.06%	73,105	40.2%
302	6,372	3.51%	79,477	43.7%
304	112	0.06%	79,589	43.8%
305	6,494	3.57%	86,083	47.4%
307	108	0.06%	86,191	47.4%
308	6,577	3.62%	92,768	51.0%
310	109	0.06%	92,877	51.1%
311	6,682	3.68%	99,559	54.8%
313	106	0.06%	99,665	54.8%
314	6,858	3.77%	106,523	58.6%
316	122	0.07%	106,645	58.7%
319	123	0.07%	106,768	58.7%
320	6,859	3.77%	113,627	62.5%
321	6,872	3.78%	120,499	66.3%
324	6,729	3.70%	127,228	70.0%
325	122	0.07%	127,350	70.1%
326	112	0.06%	127,462	70.1%
327	6,696	3.68%	134,158	73.8%
329	107	0.06%	134,265	73.9%
331	6,498	3.57%	140,763	77.4%
332	99	0.05%	140,862	77.5%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
334	6,258	3.44%	147,120	80.9%
336	85	0.05%	147,205	81.0%
338	5,708	3.14%	152,913	84.1%
339	87	0.05%	153,000	84.2%
343	5,456	3.00%	158,456	87.2%
346	4,916	2.70%	163,372	89.9%
348	62	0.03%	163,434	89.9%
351	4,498	2.47%	167,932	92.4%
356	3,781	2.08%	171,713	94.5%
361	3,094	1.70%	174,807	96.2%
366	26	0.01%	174,833	96.2%
367	2,401	1.32%	177,234	97.5%
372	19	0.01%	177,253	97.5%
374	1,858	1.02%	179,111	98.5%
379	18	0.01%	179,129	98.5%
382	1,268	0.70%	180,397	99.2%
387	7	0.00%	180,404	99.2%
392	749	0.41%	181,153	99.7%
397	3	0.00%	181,156	99.7%
400	405	0.22%	181,561	99.9%
408	183	0.10%	181,744	100%
416	43	0.02%	181,787	100%

Table Q27. PBT ELA Grade 5 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
126	8	0.00%	8	0.00%
134	12	0.01%	20	0.01%
142	16	0.01%	36	0.02%
144	2	0.00%	38	0.02%
150	22	0.01%	60	0.04%
158	63	0.04%	123	0.07%
166	114	0.07%	237	0.14%
168	2	0.00%	239	0.14%
174	217	0.13%	456	0.27%
176	6	0.00%	462	0.27%
182	366	0.21%	828	0.49%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
184	9	0.01%	837	0.49%
190	519	0.30%	1,356	0.80%
192	10	0.01%	1,366	0.80%
198	787	0.46%	2,153	1.26%
200	13	0.01%	2,166	1.27%
206	975	0.57%	3,141	1.84%
208	14	0.01%	3,155	1.85%
214	1,234	0.72%	4,389	2.57%
216	20	0.01%	4,409	2.58%
222	1,486	0.87%	5,895	3.46%
224	34	0.02%	5,929	3.48%
229	1,737	1.02%	7,666	4.49%
231	37	0.02%	7,703	4.52%
234	2,018	1.18%	9,721	5.70%
236	27	0.02%	9,748	5.72%
240	2,296	1.35%	12,044	7.06%
242	29	0.02%	12,073	7.08%
245	2,499	1.47%	14,572	8.54%
247	44	0.03%	14,616	8.57%
249	2,727	1.60%	17,343	10.2%
251	50	0.03%	17,393	10.2%
253	2,907	1.70%	20,300	11.9%
255	56	0.03%	20,356	11.9%
257	3,146	1.84%	23,502	13.8%
259	51	0.03%	23,553	13.8%
261	3,308	1.94%	26,861	15.7%
263	73	0.04%	26,934	15.8%
265	3,533	2.07%	30,467	17.9%
267	58	0.03%	30,525	17.9%
268	3,729	2.19%	34,254	20.1%
270	72	0.04%	34,326	20.1%
272	3,890	2.28%	38,216	22.4%
274	70	0.04%	38,286	22.4%
275	4,149	2.43%	42,435	24.9%
277	66	0.04%	42,501	24.9%
279	4,321	2.53%	46,822	27.5%
281	81	0.05%	46,903	27.5%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
282	4,509	2.64%	51,412	30.1%
284	71	0.04%	51,483	30.2%
285	4,618	2.71%	56,101	32.9%
287	92	0.05%	56,193	32.9%
289	4,896	2.87%	61,089	35.8%
291	4,919	2.88%	66,008	38.7%
293	79	0.05%	66,087	38.7%
295	5,162	3.03%	71,249	41.8%
297	85	0.05%	71,334	41.8%
298	5,277	3.09%	76,611	44.9%
300	75	0.04%	76,686	45.0%
301	5,374	3.15%	82,060	48.1%
303	102	0.06%	82,162	48.2%
304	5,346	3.13%	87,508	51.3%
306	94	0.06%	87,602	51.4%
307	5,509	3.23%	93,111	54.6%
309	88	0.05%	93,199	54.6%
310	5,737	3.36%	98,936	58.0%
312	79	0.05%	99,015	58.1%
314	5,586	3.28%	104,601	61.3%
316	82	0.05%	104,683	61.4%
317	5,664	3.32%	110,347	64.7%
319	94	0.06%	110,441	64.8%
320	5,735	3.36%	116,176	68.1%
322	87	0.05%	116,263	68.2%
323	5,721	3.35%	121,984	71.5%
325	81	0.05%	122,065	71.6%
327	5,575	3.27%	127,640	74.8%
329	96	0.06%	127,736	74.9%
330	5,534	3.24%	133,270	78.1%
332	65	0.04%	133,335	78.2%
334	5,221	3.06%	138,556	81.2%
336	67	0.04%	138,623	81.3%
338	5,031	2.95%	143,654	84.2%
340	65	0.04%	143,719	84.3%
342	4,738	2.78%	148,457	87.0%
344	65	0.04%	148,522	87.1%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
346	4,347	2.55%	152,869	89.6%
348	53	0.03%	152,922	89.7%
350	3,844	2.25%	156,766	91.9%
352	36	0.02%	156,802	91.9%
355	3,414	2.00%	160,216	93.9%
357	39	0.02%	160,255	94.0%
360	2,827	1.66%	163,082	95.6%
362	24	0.01%	163,106	95.6%
365	2,424	1.42%	165,530	97.0%
367	18	0.01%	165,548	97.1%
371	1,815	1.06%	167,363	98.1%
373	16	0.01%	167,379	98.1%
378	1,307	0.77%	168,686	98.9%
380	7	0.00%	168,693	98.9%
386	854	0.50%	169,547	99.4%
388	7	0.00%	169,554	99.4%
396	543	0.32%	170,097	99.7%
398	3	0.00%	170,100	99.7%
404	294	0.17%	170,394	99.9%
412	129	0.08%	170,523	100%
420	32	0.02%	170,555	100%
428	9	0.01%	170,564	100%

Table Q28. PBT ELA Grade 6 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
128	6	0.00%	6	0.00%
136	15	0.01%	21	0.01%
144	17	0.01%	38	0.02%
152	26	0.02%	64	0.04%
157	1	0.00%	65	0.04%
160	40	0.02%	105	0.06%
165	1	0.00%	106	0.06%
168	106	0.06%	212	0.13%
173	3	0.00%	215	0.13%
176	191	0.11%	406	0.24%
181	7	0.00%	413	0.25%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
184	343	0.21%	756	0.45%
189	2	0.00%	758	0.45%
192	478	0.29%	1,236	0.74%
197	11	0.01%	1,247	0.75%
201	665	0.40%	1,912	1.14%
206	11	0.01%	1,923	1.15%
209	836	0.50%	2,759	1.65%
214	18	0.01%	2,777	1.66%
217	1,082	0.65%	3,859	2.31%
222	17	0.01%	3,876	2.32%
225	1,238	0.74%	5,114	3.06%
230	1,514	0.91%	6,628	3.96%
234	1,758	1.05%	8,386	5.02%
235	24	0.01%	8,410	5.03%
239	1,871	1.12%	10,281	6.15%
243	2,021	1.21%	12,302	7.36%
244	26	0.02%	12,328	7.37%
246	2,184	1.31%	14,512	8.68%
248	33	0.02%	14,545	8.70%
250	2,313	1.38%	16,858	10.1%
251	30	0.02%	16,888	10.1%
253	2,565	1.53%	19,453	11.6%
255	35	0.02%	19,488	11.7%
257	2,618	1.57%	22,106	13.2%
258	41	0.02%	22,147	13.2%
260	2,862	1.71%	25,009	15.0%
262	39	0.02%	25,048	15.0%
263	2,912	1.74%	27,960	16.7%
265	54	0.03%	28,014	16.8%
266	3,039	1.82%	31,053	18.6%
268	37	0.02%	31,090	18.6%
269	3,173	1.90%	34,263	20.5%
271	60	0.04%	34,323	20.5%
272	3,327	1.99%	37,650	22.5%
274	51	0.03%	37,701	22.6%
275	3,427	2.05%	41,128	24.6%
277	3,510	2.10%	44,638	26.7%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
280	3,641	2.18%	48,279	28.9%
282	61	0.04%	48,340	28.9%
283	3,894	2.33%	52,234	31.2%
285	3,956	2.37%	56,190	33.6%
288	4,040	2.42%	60,230	36.0%
290	4,203	2.51%	64,433	38.5%
293	4,356	2.61%	68,789	41.1%
295	4,556	2.73%	73,345	43.9%
298	4,613	2.76%	77,958	46.6%
300	4,582	2.74%	82,540	49.4%
303	4,802	2.87%	87,342	52.2%
305	4,837	2.89%	92,179	55.1%
308	4,942	2.96%	97,121	58.1%
310	5,144	3.08%	102,265	61.2%
313	5,300	3.17%	107,565	64.3%
315	95	0.06%	107,660	64.4%
316	5,409	3.24%	113,069	67.6%
318	89	0.05%	113,158	67.7%
320	5,241	3.13%	118,399	70.8%
321	105	0.06%	118,504	70.9%
322	5,460	3.27%	123,964	74.2%
325	5,426	3.25%	129,390	77.4%
327	87	0.05%	129,477	77.4%
328	5,366	3.21%	134,843	80.7%
330	85	0.05%	134,928	80.7%
332	5,314	3.18%	140,242	83.9%
333	91	0.05%	140,333	83.9%
337	98	0.06%	140,431	84.0%
338	5,008	3.00%	145,439	87.0%
340	4,877	2.92%	150,316	89.9%
343	51	0.03%	150,367	89.9%
344	4,456	2.67%	154,823	92.6%
345	74	0.04%	154,897	92.7%
349	48	0.03%	154,945	92.7%
350	3,745	2.24%	158,690	94.9%
355	38	0.02%	158,728	94.9%
356	3,109	1.86%	161,837	96.8%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
361	22	0.01%	161,859	96.8%
363	2,297	1.37%	164,156	98.2%
368	22	0.01%	164,178	98.2%
373	1,627	0.97%	165,805	99.2%
378	12	0.01%	165,817	99.2%
387	895	0.54%	166,712	99.7%
392	4	0.00%	166,716	99.7%
395	374	0.22%	167,090	99.9%
400	3	0.00%	167,093	99.9%
403	87	0.05%	167,180	100%

Table Q29. PBT ELA Grade 7 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
133	8	0.01%	8	0.01%
140	17	0.01%	25	0.02%
148	20	0.01%	45	0.03%
156	24	0.02%	69	0.04%
164	67	0.04%	136	0.09%
166	2	0.00%	138	0.09%
172	80	0.05%	218	0.14%
174	2	0.00%	220	0.14%
180	140	0.09%	360	0.23%
182	3	0.00%	363	0.23%
188	256	0.16%	619	0.39%
190	5	0.00%	624	0.40%
196	339	0.22%	963	0.61%
198	6	0.00%	969	0.62%
204	470	0.30%	1,439	0.92%
206	13	0.01%	1,452	0.92%
212	647	0.41%	2,099	1.34%
214	17	0.01%	2,116	1.35%
220	753	0.48%	2,869	1.83%
222	19	0.01%	2,888	1.84%
228	929	0.59%	3,817	2.43%
230	28	0.02%	3,845	2.45%
235	1,070	0.68%	4,915	3.13%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
237	31	0.02%	4,946	3.15%
240	1,214	0.77%	6,160	3.92%
242	33	0.02%	6,193	3.94%
245	1,324	0.84%	7,517	4.78%
247	37	0.02%	7,554	4.81%
249	1,585	1.01%	9,139	5.81%
251	32	0.02%	9,171	5.83%
253	1,659	1.06%	10,830	6.89%
255	35	0.02%	10,865	6.91%
257	1,840	1.17%	12,705	8.08%
259	44	0.03%	12,749	8.11%
261	2,009	1.28%	14,758	9.39%
263	67	0.04%	14,825	9.43%
264	2,172	1.38%	16,997	10.8%
266	48	0.03%	17,045	10.8%
268	2,349	1.49%	19,394	12.3%
270	61	0.04%	19,455	12.4%
271	2,465	1.57%	21,920	13.9%
273	82	0.05%	22,002	14.0%
274	2,729	1.74%	24,731	15.7%
276	60	0.04%	24,791	15.8%
277	2,914	1.85%	27,705	17.6%
279	65	0.04%	27,770	17.7%
280	3,056	1.94%	30,826	19.6%
282	63	0.04%	30,889	19.7%
283	3,169	2.02%	34,058	21.7%
285	82	0.05%	34,140	21.7%
287	3,364	2.14%	37,504	23.9%
288	3,452	2.20%	40,956	26.1%
289	74	0.05%	41,030	26.1%
290	97	0.06%	41,127	26.2%
291	3,690	2.35%	44,817	28.5%
293	3,947	2.51%	48,764	31.0%
295	93	0.06%	48,857	31.1%
296	4,042	2.57%	52,899	33.7%
298	4,355	2.77%	57,254	36.4%
300	98	0.06%	57,352	36.5%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
301	4,355	2.77%	61,707	39.3%
303	4,627	2.94%	66,334	42.2%
305	4,651	2.96%	70,985	45.2%
307	116	0.07%	71,101	45.2%
308	4,757	3.03%	75,858	48.3%
310	4,964	3.16%	80,822	51.4%
312	102	0.06%	80,924	51.5%
313	5,014	3.19%	85,938	54.7%
315	5,215	3.32%	91,153	58.0%
317	123	0.08%	91,276	58.1%
318	5,225	3.32%	96,501	61.4%
320	5,345	3.40%	101,846	64.8%
322	118	0.08%	101,964	64.9%
323	5,196	3.31%	107,160	68.2%
325	102	0.06%	107,262	68.2%
326	5,114	3.25%	112,376	71.5%
328	5,158	3.28%	117,534	74.8%
330	99	0.06%	117,633	74.8%
331	5,104	3.25%	122,737	78.1%
333	115	0.07%	122,852	78.2%
334	4,886	3.11%	127,738	81.3%
336	83	0.05%	127,821	81.3%
338	4,679	2.98%	132,500	84.3%
340	95	0.06%	132,595	84.4%
341	4,428	2.82%	137,023	87.2%
343	76	0.05%	137,099	87.2%
347	4,136	2.63%	141,235	89.9%
349	3,974	2.53%	145,209	92.4%
351	58	0.04%	145,267	92.4%
354	3,365	2.14%	148,632	94.6%
356	66	0.04%	148,698	94.6%
359	2,906	1.85%	151,604	96.5%
361	39	0.02%	151,643	96.5%
366	2,268	1.44%	153,911	97.9%
368	37	0.02%	153,948	97.9%
374	1,607	1.02%	155,555	99.0%
376	19	0.01%	155,574	99.0%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
386	1,018	0.65%	156,592	99.6%
388	11	0.01%	156,603	99.6%
394	467	0.30%	157,070	99.9%
396	2	0.00%	157,072	99.9%
402	110	0.07%	157,182	100%

Table Q30. PBT ELA Grade 8 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
121	12	0.01%	12	0.01%
129	23	0.02%	35	0.02%
137	24	0.02%	59	0.04%
145	21	0.01%	80	0.05%
153	37	0.02%	117	0.08%
161	62	0.04%	179	0.12%
168	143	0.10%	322	0.22%
176	198	0.13%	520	0.35%
184	287	0.19%	807	0.54%
192	359	0.24%	1,166	0.78%
200	478	0.32%	1,644	1.10%
208	541	0.36%	2,185	1.46%
216	642	0.43%	2,827	1.90%
224	745	0.50%	3,572	2.39%
229	898	0.60%	4,470	3.00%
234	999	0.67%	5,469	3.67%
239	1,103	0.74%	6,572	4.41%
243	1,234	0.83%	7,806	5.23%
247	1,337	0.90%	9,143	6.13%
251	1,444	0.97%	10,587	7.10%
254	1,547	1.04%	12,134	8.14%
257	1,589	1.07%	13,723	9.20%
261	1,719	1.15%	15,442	10.4%
264	1,870	1.25%	17,312	11.6%
267	2,068	1.39%	19,380	13.0%
270	2,089	1.40%	21,469	14.4%
273	2,272	1.52%	23,741	15.9%
275	2,421	1.62%	26,162	17.5%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
278	2,573	1.73%	28,735	19.3%
281	2,742	1.84%	31,477	21.1%
284	2,961	1.99%	34,438	23.1%
286	3,170	2.13%	37,608	25.2%
288	3,364	2.26%	40,972	27.5%
291	3,502	2.35%	44,474	29.8%
293	3,868	2.59%	48,342	32.4%
296	4,106	2.75%	52,448	35.2%
298	4,233	2.84%	56,681	38.0%
301	4,338	2.91%	61,019	40.9%
304	4,763	3.19%	65,782	44.1%
306	4,918	3.30%	70,700	47.4%
309	5,251	3.52%	75,951	50.9%
311	5,216	3.50%	81,167	54.4%
316	5,531	3.71%	86,698	58.1%
317	5,696	3.82%	92,394	61.9%
320	5,627	3.77%	98,021	65.7%
323	5,947	3.99%	103,968	69.7%
326	5,923	3.97%	109,891	73.7%
330	5,627	3.77%	115,518	77.5%
334	5,658	3.79%	121,176	81.2%
338	5,640	3.78%	126,816	85.0%
343	5,261	3.53%	132,077	88.6%
347	4,701	3.15%	136,778	91.7%
354	4,046	2.71%	140,824	94.4%
361	3,320	2.23%	144,144	96.6%
371	2,526	1.69%	146,670	98.3%
386	1,550	1.04%	148,220	99.4%
394	752	0.50%	148,972	99.9%
402	176	0.12%	149,148	100%

Table Q31. PBT Mathematics Grade 3 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
145	6	0.00%	6	0.00%
153	13	0.01%	19	0.01%
161	16	0.01%	35	0.02%
169	58	0.03%	93	0.05%
177	136	0.07%	229	0.12%
185	265	0.14%	494	0.27%
189	1	0.00%	495	0.27%
193	508	0.28%	1,003	0.55%
197	7	0.00%	1,010	0.55%
201	812	0.44%	1,822	0.99%
205	7	0.00%	1,829	1.00%
209	1,162	0.63%	2,991	1.63%
213	20	0.01%	3,011	1.64%
217	1,639	0.89%	4,650	2.53%
221	19	0.01%	4,669	2.54%
225	1,976	1.08%	6,645	3.62%
229	26	0.01%	6,671	3.63%
233	2,327	1.27%	8,998	4.90%
237	25	0.01%	9,023	4.92%
239	2,494	1.36%	11,517	6.28%
243	26	0.01%	11,543	6.29%
245	2,669	1.45%	14,212	7.74%
249	33	0.02%	14,245	7.76%
250	2,817	1.53%	17,062	9.30%
254	2,979	1.62%	20,041	10.9%
258	2,892	1.58%	22,933	12.5%
262	3,064	1.67%	25,997	14.2%
266	3,091	1.68%	29,088	15.8%
269	3,107	1.69%	32,195	17.5%
270	33	0.02%	32,228	17.6%
272	3,292	1.79%	35,520	19.4%
273	37	0.02%	35,557	19.4%
275	3,268	1.78%	38,825	21.2%
276	55	0.03%	38,880	21.2%
278	3,422	1.86%	42,302	23.0%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
279	37	0.02%	42,339	23.1%
281	3,400	1.85%	45,739	24.9%
282	57	0.03%	45,796	25.0%
285	3,510	1.91%	49,306	26.9%
286	3,365	1.83%	52,671	28.7%
288	3,519	1.92%	56,190	30.6%
289	46	0.03%	56,236	30.6%
290	3,565	1.94%	59,801	32.6%
292	55	0.03%	59,856	32.6%
293	3,704	2.02%	63,560	34.6%
294	54	0.03%	63,614	34.7%
295	3,763	2.05%	67,377	36.7%
297	3,790	2.07%	71,167	38.8%
299	3,787	2.06%	74,954	40.8%
301	56	0.03%	75,010	40.9%
302	3,878	2.11%	78,888	43.0%
303	66	0.04%	78,954	43.0%
304	3,972	2.16%	82,926	45.2%
306	4,053	2.21%	86,979	47.4%
308	4,293	2.34%	91,272	49.7%
310	4,086	2.23%	95,358	52.0%
312	65	0.04%	95,423	52.0%
314	4,253	2.32%	99,676	54.3%
315	4,263	2.32%	103,939	56.6%
317	4,249	2.32%	108,188	58.9%
318	86	0.05%	108,274	59.0%
319	4,385	2.39%	112,659	61.4%
321	4,511	2.46%	117,170	63.8%
323	86	0.05%	117,256	63.9%
324	4,613	2.51%	121,869	66.4%
325	71	0.04%	121,940	66.4%
326	4,638	2.53%	126,578	69.0%
328	81	0.04%	126,659	69.0%
329	4,713	2.57%	131,372	71.6%
330	76	0.04%	131,448	71.6%
332	4,791	2.61%	136,239	74.2%
333	52	0.03%	136,291	74.3%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
335	4,838	2.64%	141,129	76.9%
336	71	0.04%	141,200	76.9%
339	70	0.04%	141,270	77.0%
340	4,912	2.68%	146,182	79.6%
342	4,934	2.69%	151,116	82.3%
344	62	0.03%	151,178	82.4%
346	5,045	2.75%	156,223	85.1%
350	5,005	2.73%	161,228	87.8%
354	63	0.03%	161,291	87.9%
355	4,855	2.65%	166,146	90.5%
359	50	0.03%	166,196	90.6%
362	4,657	2.54%	170,853	93.1%
366	58	0.03%	170,911	93.1%
370	4,254	2.32%	175,165	95.4%
374	33	0.02%	175,198	95.5%
381	3,726	2.03%	178,924	97.5%
385	29	0.02%	178,953	97.5%
389	2,898	1.58%	181,851	99.1%
393	31	0.02%	181,882	99.1%
397	1,651	0.90%	183,533	100%

Table Q32. PBT Mathematics Grade 4 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
133	10	0.01%	10	0.01%
141	10	0.01%	20	0.01%
149	18	0.01%	38	0.02%
157	45	0.02%	83	0.05%
165	106	0.06%	189	0.10%
173	219	0.12%	408	0.22%
178	1	0.00%	409	0.22%
182	416	0.23%	825	0.45%
187	2	0.00%	827	0.45%
190	705	0.38%	1,532	0.83%
195	3	0.00%	1,535	0.84%
198	1,062	0.58%	2,597	1.41%
203	5	0.00%	2,602	1.42%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
206	1,362	0.74%	3,964	2.16%
211	12	0.01%	3,976	2.17%
214	1,732	0.94%	5,708	3.11%
219	14	0.01%	5,722	3.12%
222	1,921	1.05%	7,643	4.16%
227	15	0.01%	7,658	4.17%
230	2,174	1.18%	9,832	5.36%
235	11	0.01%	9,843	5.36%
236	2,311	1.26%	12,154	6.62%
241	2,398	1.31%	14,552	7.93%
245	2,535	1.38%	17,087	9.31%
246	16	0.01%	17,103	9.32%
250	2,538	1.38%	19,641	10.7%
254	2,684	1.46%	22,325	12.2%
255	16	0.01%	22,341	12.2%
257	2,792	1.52%	25,133	13.7%
259	20	0.01%	25,153	13.7%
261	2,732	1.49%	27,885	15.2%
262	21	0.01%	27,906	15.2%
264	2,887	1.57%	30,793	16.8%
266	28	0.02%	30,821	16.8%
267	2,961	1.61%	33,782	18.4%
269	3,012	1.64%	36,794	20.0%
272	3,140	1.71%	39,934	21.8%
274	36	0.02%	39,970	21.8%
275	3,106	1.69%	43,076	23.5%
277	3,211	1.75%	46,287	25.2%
279	3,177	1.73%	49,464	26.9%
280	35	0.02%	49,499	27.0%
282	28	0.02%	49,527	27.0%
283	3,254	1.77%	52,781	28.8%
284	3,316	1.81%	56,097	30.6%
286	3,179	1.73%	59,276	32.3%
288	3,352	1.83%	62,628	34.1%
289	28	0.02%	62,656	34.1%
290	3,391	1.85%	66,047	36.0%
291	22	0.01%	66,069	36.0%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
292	3,440	1.87%	69,509	37.9%
293	43	0.02%	69,552	37.9%
294	3,490	1.90%	73,042	39.8%
295	33	0.02%	73,075	39.8%
296	3,413	1.86%	76,488	41.7%
297	22	0.01%	76,510	41.7%
298	3,469	1.89%	79,979	43.6%
299	35	0.02%	80,014	43.6%
300	3,497	1.91%	83,511	45.5%
301	34	0.02%	83,545	45.5%
302	3,583	1.95%	87,128	47.5%
303	37	0.02%	87,165	47.5%
304	3,674	2.00%	90,839	49.5%
305	40	0.02%	90,879	49.5%
306	3,666	2.00%	94,545	51.5%
307	36	0.02%	94,581	51.5%
308	3,601	1.96%	98,182	53.5%
309	40	0.02%	98,222	53.5%
310	3,691	2.01%	101,913	55.5%
311	43	0.02%	101,956	55.5%
312	3,731	2.03%	105,687	57.6%
313	38	0.02%	105,725	57.6%
314	3,674	2.00%	109,399	59.6%
315	42	0.02%	109,441	59.6%
316	3,866	2.11%	113,307	61.7%
317	39	0.02%	113,346	61.8%
319	3,955	2.15%	117,301	63.9%
321	3,894	2.12%	121,195	66.0%
323	4,103	2.24%	125,298	68.3%
324	35	0.02%	125,333	68.3%
325	4,046	2.20%	129,379	70.5%
326	38	0.02%	129,417	70.5%
328	4,124	2.25%	133,541	72.8%
330	4,110	2.24%	137,651	75.0%
333	4,057	2.21%	141,708	77.2%
335	45	0.02%	141,753	77.2%
336	4,161	2.27%	145,914	79.5%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
338	37	0.02%	145,951	79.5%
341	4,250	2.32%	150,201	81.8%
343	4,334	2.36%	154,535	84.2%
346	39	0.02%	154,574	84.2%
347	4,315	2.35%	158,889	86.6%
348	19	0.01%	158,908	86.6%
351	4,420	2.41%	163,328	89.0%
352	34	0.02%	163,362	89.0%
356	4,260	2.32%	167,622	91.3%
361	26	0.01%	167,648	91.3%
362	4,118	2.24%	171,766	93.6%
367	26	0.01%	171,792	93.6%
370	3,877	2.11%	175,669	95.7%
375	14	0.01%	175,683	95.7%
381	3,398	1.85%	179,081	97.6%
386	18	0.01%	179,099	97.6%
389	2,757	1.50%	181,856	99.1%
394	7	0.00%	181,863	99.1%
397	1,690	0.92%	183,553	100%

Table Q33. PBT Mathematics Grade 5 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
151	6	0.00%	6	0.00%
159	4	0.00%	10	0.01%
167	24	0.01%	34	0.02%
169	1	0.00%	35	0.02%
175	63	0.04%	98	0.06%
177	3	0.00%	101	0.06%
183	176	0.10%	277	0.16%
185	1	0.00%	278	0.16%
191	416	0.24%	694	0.40%
193	3	0.00%	697	0.41%
199	722	0.42%	1,419	0.83%
201	9	0.01%	1,428	0.83%
207	1,218	0.71%	2,646	1.54%
209	13	0.01%	2,659	1.55%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
215	1,696	0.99%	4,355	2.54%
217	18	0.01%	4,373	2.55%
223	2,277	1.33%	6,650	3.88%
225	19	0.01%	6,669	3.89%
232	2,744	1.60%	9,413	5.49%
234	23	0.01%	9,436	5.50%
242	3,012	1.76%	12,448	7.26%
244	39	0.02%	12,487	7.28%
250	3,311	1.93%	15,798	9.21%
252	26	0.02%	15,824	9.23%
256	3,247	1.89%	19,071	11.1%
258	32	0.02%	19,103	11.1%
261	3,364	1.96%	22,467	13.1%
263	33	0.02%	22,500	13.1%
265	3,323	1.94%	25,823	15.1%
267	30	0.02%	25,853	15.1%
269	3,289	1.92%	29,142	17.0%
271	24	0.01%	29,166	17.0%
273	3,238	1.89%	32,404	18.9%
275	33	0.02%	32,437	18.9%
276	3,368	1.96%	35,805	20.9%
278	31	0.02%	35,836	20.9%
279	3,285	1.92%	39,121	22.8%
281	31	0.02%	39,152	22.8%
282	3,241	1.89%	42,393	24.7%
284	3,327	1.94%	45,720	26.7%
286	40	0.02%	45,760	26.7%
287	3,274	1.91%	49,034	28.6%
289	3,337	1.95%	52,371	30.5%
291	3,288	1.92%	55,659	32.5%
293	36	0.02%	55,695	32.5%
294	3,264	1.90%	58,959	34.4%
296	3,338	1.95%	62,297	36.3%
298	3,211	1.87%	65,508	38.2%
300	3,320	1.94%	68,828	40.1%
302	3,355	1.96%	72,183	42.1%
304	3,313	1.93%	75,496	44.0%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
305	3,327	1.94%	78,823	46.0%
306	56	0.03%	78,879	46.0%
307	3,224	1.88%	82,103	47.9%
309	3,192	1.86%	85,295	49.8%
311	3,266	1.91%	88,561	51.7%
313	3,262	1.90%	91,823	53.6%
314	3,183	1.86%	95,006	55.4%
315	41	0.02%	95,047	55.4%
316	3,315	1.93%	98,362	57.4%
318	45	0.03%	98,407	57.4%
319	3,239	1.89%	101,646	59.3%
320	3,335	1.95%	104,981	61.2%
321	3,265	1.90%	108,246	63.1%
322	38	0.02%	108,284	63.2%
323	3,350	1.95%	111,634	65.1%
325	3,325	1.94%	114,959	67.1%
327	3,311	1.93%	118,270	69.0%
329	3,441	2.01%	121,711	71.0%
331	3,453	2.01%	125,164	73.0%
332	3,310	1.93%	128,474	74.9%
333	41	0.02%	128,515	75.0%
334	3,234	1.89%	131,749	76.8%
336	29	0.02%	131,778	76.9%
337	3,272	1.91%	135,050	78.8%
339	3,240	1.89%	138,290	80.7%
341	3,304	1.93%	141,594	82.6%
343	3,214	1.87%	144,808	84.5%
345	29	0.02%	144,837	84.5%
346	3,223	1.88%	148,060	86.4%
348	27	0.02%	148,087	86.4%
349	2,952	1.72%	151,039	88.1%
351	18	0.01%	151,057	88.1%
352	3,007	1.75%	154,064	89.9%
354	19	0.01%	154,083	89.9%
355	2,897	1.69%	156,980	91.6%
357	22	0.01%	157,002	91.6%
359	2,871	1.67%	159,873	93.3%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
361	23	0.01%	159,896	93.3%
363	2,672	1.56%	162,568	94.8%
365	21	0.01%	162,589	94.8%
369	2,400	1.40%	164,989	96.2%
371	17	0.01%	165,006	96.2%
375	2,152	1.26%	167,158	97.5%
377	12	0.01%	167,170	97.5%
385	1,868	1.09%	169,038	98.6%
387	9	0.01%	169,047	98.6%
393	1,512	0.88%	170,559	99.5%
395	8	0.00%	170,567	99.5%
401	876	0.51%	171,443	100%

Table Q34. PBT Mathematics Grade 6 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
132	12	0.01%	12	0.01%
140	10	0.01%	22	0.01%
148	22	0.01%	44	0.03%
151	1	0.00%	45	0.03%
157	37	0.02%	82	0.05%
160	1	0.00%	83	0.05%
165	71	0.04%	154	0.09%
168	1	0.00%	155	0.09%
173	172	0.10%	327	0.20%
176	1	0.00%	328	0.20%
181	353	0.21%	681	0.41%
184	5	0.00%	686	0.41%
189	574	0.34%	1,260	0.75%
192	8	0.00%	1,268	0.76%
197	970	0.58%	2,238	1.34%
200	16	0.01%	2,254	1.35%
205	1,476	0.88%	3,730	2.23%
208	12	0.01%	3,742	2.24%
213	1,991	1.19%	5,733	3.43%
216	18	0.01%	5,751	3.44%
221	2,477	1.48%	8,228	4.93%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
224	22	0.01%	8,250	4.94%
230	2,906	1.74%	11,156	6.68%
233	29	0.02%	11,185	6.70%
238	3,234	1.94%	14,419	8.63%
241	32	0.02%	14,451	8.65%
245	3,458	2.07%	17,909	10.7%
248	34	0.02%	17,943	10.7%
251	3,642	2.18%	21,585	12.9%
254	34	0.02%	21,619	12.9%
256	3,620	2.17%	25,239	15.1%
259	40	0.02%	25,279	15.1%
261	3,632	2.17%	28,911	17.3%
264	34	0.02%	28,945	17.3%
265	3,603	2.16%	32,548	19.5%
268	3,646	2.18%	36,194	21.7%
271	38	0.02%	36,232	21.7%
272	3,459	2.07%	39,691	23.8%
275	3,504	2.10%	43,195	25.9%
278	3,504	2.10%	46,699	28.0%
281	3,344	2.00%	50,043	30.0%
284	3,358	2.01%	53,401	32.0%
286	3,316	1.99%	56,717	34.0%
287	48	0.03%	56,765	34.0%
289	3,418	2.05%	60,183	36.0%
291	3,267	1.96%	63,450	38.0%
292	47	0.03%	63,497	38.0%
293	3,257	1.95%	66,754	40.0%
294	52	0.03%	66,806	40.0%
295	3,107	1.86%	69,913	41.9%
296	48	0.03%	69,961	41.9%
298	3,249	1.95%	73,210	43.8%
300	3,208	1.92%	76,418	45.7%
301	35	0.02%	76,453	45.8%
302	3,196	1.91%	79,649	47.7%
303	50	0.03%	79,699	47.7%
304	3,135	1.88%	82,834	49.6%
305	44	0.03%	82,878	49.6%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
306	3,043	1.82%	85,921	51.4%
307	59	0.04%	85,980	51.5%
308	3,121	1.87%	89,101	53.3%
309	52	0.03%	89,153	53.4%
310	3,059	1.83%	92,212	55.2%
311	49	0.03%	92,261	55.2%
312	2,988	1.79%	95,249	57.0%
313	33	0.02%	95,282	57.0%
314	2,937	1.76%	98,219	58.8%
315	59	0.04%	98,278	58.8%
316	2,955	1.77%	101,233	60.6%
317	51	0.03%	101,284	60.6%
318	2,921	1.75%	104,205	62.4%
319	2,933	1.76%	107,138	64.1%
321	2,962	1.77%	110,100	65.9%
322	48	0.03%	110,148	65.9%
323	2,815	1.69%	112,963	67.6%
324	53	0.03%	113,016	67.7%
325	2,806	1.68%	115,822	69.3%
326	50	0.03%	115,872	69.4%
327	2,786	1.67%	118,658	71.0%
328	60	0.04%	118,718	71.1%
329	2,802	1.68%	121,520	72.8%
330	49	0.03%	121,569	72.8%
331	2,769	1.66%	124,338	74.4%
332	52	0.03%	124,390	74.5%
333	2,745	1.64%	127,135	76.1%
334	50	0.03%	127,185	76.1%
335	2,730	1.63%	129,915	77.8%
336	48	0.03%	129,963	77.8%
337	2,649	1.59%	132,612	79.4%
338	48	0.03%	132,660	79.4%
340	2,626	1.57%	135,286	81.0%
342	2,644	1.58%	137,930	82.6%
343	43	0.03%	137,973	82.6%
344	2,510	1.50%	140,483	84.1%
345	36	0.02%	140,519	84.1%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
347	2,686	1.61%	143,205	85.7%
349	2,553	1.53%	145,758	87.3%
350	39	0.02%	145,797	87.3%
352	2,484	1.49%	148,281	88.8%
355	2,489	1.49%	150,770	90.3%
358	2,430	1.45%	153,200	91.7%
361	45	0.03%	153,245	91.7%
362	2,236	1.34%	155,481	93.1%
365	32	0.02%	155,513	93.1%
366	2,157	1.29%	157,670	94.4%
369	26	0.02%	157,696	94.4%
370	2,046	1.22%	159,742	95.6%
373	21	0.01%	159,763	95.6%
375	1,795	1.07%	161,558	96.7%
378	19	0.01%	161,577	96.7%
381	1,547	0.93%	163,124	97.7%
384	13	0.01%	163,137	97.7%
388	1,348	0.81%	164,485	98.5%
391	8	0.00%	164,493	98.5%
397	1,137	0.68%	165,630	99.2%
400	9	0.01%	165,639	99.2%
405	794	0.48%	166,433	99.6%
408	5	0.00%	166,438	99.6%
413	486	0.29%	166,924	99.9%
416	1	0.00%	166,925	99.9%
421	109	0.07%	167,034	100%

Table Q35. PBT Mathematics Grade 7 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
160	15	0.01%	15	0.01%
168	14	0.01%	29	0.02%
176	40	0.03%	69	0.04%
184	87	0.06%	156	0.10%
191	199	0.13%	355	0.23%
199	430	0.28%	785	0.51%
207	804	0.52%	1,589	1.02%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
215	1,425	0.92%	3,014	1.94%
223	2,196	1.41%	5,210	3.36%
230	3,085	1.99%	8,295	5.34%
238	3,847	2.48%	12,142	7.82%
246	4,560	2.94%	16,702	10.8%
259	4,756	3.06%	21,458	13.8%
267	4,875	3.14%	26,333	17.0%
273	4,729	3.05%	31,062	20.0%
277	4,467	2.88%	35,529	22.9%
281	4,285	2.76%	39,814	25.6%
285	4,001	2.58%	43,815	28.2%
288	3,794	2.44%	47,609	30.7%
291	3,592	2.31%	51,201	33.0%
293	3,368	2.17%	54,569	35.1%
295	3,317	2.14%	57,886	37.3%
297	3,100	2.00%	60,986	39.3%
299	2,930	1.89%	63,916	41.2%
301	2,890	1.86%	66,806	43.0%
303	2,651	1.71%	69,457	44.7%
305	2,664	1.72%	72,121	46.5%
306	2,515	1.62%	74,636	48.1%
308	2,399	1.55%	77,035	49.6%
309	2,440	1.57%	79,475	51.2%
311	2,273	1.46%	81,748	52.7%
312	2,245	1.45%	83,993	54.1%
314	2,236	1.44%	86,229	55.5%
315	2,298	1.48%	88,527	57.0%
317	2,184	1.41%	90,711	58.4%
318	2,173	1.40%	92,884	59.8%
319	2,140	1.38%	95,024	61.2%
321	2,160	1.39%	97,184	62.6%
322	2,168	1.40%	99,352	64.0%
323	2,105	1.36%	101,457	65.3%
325	2,120	1.37%	103,577	66.7%
326	2,150	1.38%	105,727	68.1%
327	2,053	1.32%	107,780	69.4%
329	2,013	1.30%	109,793	70.7%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
330	2,004	1.29%	111,797	72.0%
331	2,070	1.33%	113,867	73.3%
333	2,036	1.31%	115,903	74.7%
334	2,055	1.32%	117,958	76.0%
335	2,021	1.30%	119,979	77.3%
337	2,071	1.33%	122,050	78.6%
338	2,027	1.31%	124,077	79.9%
340	2,047	1.32%	126,124	81.2%
341	2,122	1.37%	128,246	82.6%
343	2,100	1.35%	130,346	84.0%
345	2,001	1.29%	132,347	85.2%
346	2,109	1.36%	134,456	86.6%
348	2,043	1.32%	136,499	87.9%
350	2,017	1.30%	138,516	89.2%
352	2,030	1.31%	140,546	90.5%
354	1,986	1.28%	142,532	91.8%
357	2,042	1.32%	144,574	93.1%
360	1,898	1.22%	146,472	94.3%
363	1,755	1.13%	148,227	95.5%
367	1,702	1.10%	149,929	96.6%
371	1,557	1.00%	151,486	97.6%
377	1,363	0.88%	152,849	98.5%
385	1,158	0.75%	154,007	99.2%
393	835	0.54%	154,842	99.7%
401	413	0.27%	155,255	100%

Table Q36. PBT Mathematics Grade 8 Scale Score Frequency Distribution

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
134	27	0.02%	27	0.02%
142	23	0.02%	50	0.04%
150	22	0.02%	72	0.06%
158	52	0.04%	124	0.11%
166	89	0.08%	213	0.18%
173	195	0.17%	408	0.35%
174	2	0.00%	410	0.35%
181	361	0.31%	771	0.66%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
189	680	0.58%	1,451	1.24%
197	1,182	1.01%	2,633	2.25%
205	1,823	1.56%	4,456	3.81%
213	2,495	2.14%	6,951	5.95%
221	3,367	2.88%	10,318	8.83%
229	4,003	3.43%	14,321	12.3%
237	43	0.04%	14,364	12.3%
244	4,357	3.73%	18,721	16.0%
252	39	0.03%	18,760	16.1%
254	4,539	3.89%	23,299	19.9%
262	4,613	3.95%	27,912	23.9%
268	4,396	3.76%	32,308	27.7%
270	43	0.04%	32,351	27.7%
273	4,244	3.63%	36,595	31.3%
276	31	0.03%	36,626	31.4%
277	3,913	3.35%	40,539	34.7%
280	3,664	3.14%	44,203	37.8%
281	29	0.02%	44,232	37.9%
284	3,421	2.93%	47,653	40.8%
285	34	0.03%	47,687	40.8%
287	3,055	2.62%	50,742	43.4%
288	29	0.02%	50,771	43.5%
289	3,022	2.59%	53,793	46.0%
292	2,803	2.40%	56,596	48.4%
294	2,588	2.22%	59,184	50.7%
295	35	0.03%	59,219	50.7%
296	2,516	2.15%	61,735	52.8%
297	26	0.02%	61,761	52.9%
298	2,510	2.15%	64,271	55.0%
300	2,303	1.97%	66,574	57.0%
302	2,349	2.01%	68,923	59.0%
304	2,253	1.93%	71,176	60.9%
306	2,150	1.84%	73,326	62.8%
307	2,053	1.76%	75,379	64.5%
308	21	0.02%	75,400	64.5%
309	1,893	1.62%	77,293	66.2%
310	26	0.02%	77,319	66.2%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
311	1,861	1.59%	79,180	67.8%
312	1,851	1.58%	81,031	69.4%
314	1,853	1.59%	82,884	70.9%
315	1,848	1.58%	84,732	72.5%
317	1,714	1.47%	86,446	74.0%
318	1,655	1.42%	88,101	75.4%
319	1,576	1.35%	89,677	76.8%
320	15	0.01%	89,692	76.8%
321	1,504	1.29%	91,196	78.1%
322	1,493	1.28%	92,689	79.3%
323	18	0.02%	92,707	79.4%
324	1,353	1.16%	94,060	80.5%
325	1,449	1.24%	95,509	81.8%
326	11	0.01%	95,520	81.8%
327	1,287	1.10%	96,807	82.9%
328	1,267	1.08%	98,074	84.0%
329	16	0.01%	98,090	84.0%
330	1,162	0.99%	99,252	85.0%
331	1,231	1.05%	100,483	86.0%
332	10	0.01%	100,493	86.0%
333	1,122	0.96%	101,615	87.0%
334	1,133	0.97%	102,748	88.0%
335	11	0.01%	102,759	88.0%
336	1,085	0.93%	103,844	88.9%
337	1,109	0.95%	104,953	89.8%
338	6	0.01%	104,959	89.8%
339	1,008	0.86%	105,967	90.7%
341	927	0.79%	106,894	91.5%
342	907	0.78%	107,801	92.3%
344	928	0.79%	108,729	93.1%
345	9	0.01%	108,738	93.1%
346	789	0.68%	109,527	93.8%
347	7	0.01%	109,534	93.8%
349	762	0.65%	110,296	94.4%
350	780	0.67%	111,076	95.1%
352	4	0.00%	111,080	95.1%
353	805	0.69%	111,885	95.8%

Appendix Q: Raw-to-Scale Score and Scale Score Frequency Tables

Scale Score	Freq.	Pct.	Cumulative	
			Freq.	Pct.
354	5	0.00%	111,890	95.8%
355	724	0.62%	112,614	96.4%
357	1	0.00%	112,615	96.4%
358	696	0.60%	113,311	97.0%
361	654	0.56%	113,965	97.6%
363	3	0.00%	113,968	97.6%
365	647	0.55%	114,615	98.1%
369	3	0.00%	114,618	98.1%
370	586	0.50%	115,204	98.6%
373	2	0.00%	115,206	98.6%
376	559	0.48%	115,765	99.1%
385	491	0.42%	116,256	99.5%
392	379	0.32%	116,635	99.8%
393	2	0.00%	116,637	99.8%
400	185	0.16%	116,822	100%

Appendix R: Study of Operational Test Mode Comparability

Section R.1. Introduction

R.1.1. Overview

The 2017 administration was the first in which the New York State Education Department (NYSED) offered its operational test (OP) in a computer-based testing (CBT) environment for the Grades 3–8 New York State English Language Arts (ELA) and Mathematics Tests. The goal of this study is to detect and begin to understand differences in student performance that may be attributable to the mode in which a student tested (i.e., paper-based testing, or “PBT,” versus CBT). The main inference to be drawn is whether scores that arise from students testing in PBT and CBT are interchangeable, so the focus will be at the form- rather than item-level. This study will be repeated over the next few years, to monitor test mode comparability as more New York State students test on the CBT platform.

Table R.1.1. Unique Items Administered in Both CBT and PBT Modes: ELA

Grade	Items			
	MC	CR2	CR3	CR4
3	31	7		2
4	31	7		2
5	42	7		2
6	42	7		2
7	35	7		2
8	42	7		2

Note. The operational Mathematics test forms contain MC, 2-point CR, and 3-point CR items, while the operational ELA test forms contain MC, 2-point CR, and 4-point CR items.

Table R.1.2. Unique Items Administered in Both CBT and PBT Modes: Mathematics

Grade	Items			
	MC	CR2	CR3	CR4
3	44	5	3	
4	45	6	4	
5	46	6	4	
6	51	6	4	
7	51	6	4	
8	51	6	4	

Note. The operational Mathematics test forms contain MC, 2-point CR, and 3-point CR items, while the operational ELA test forms contain MC, 2-point CR, and 4-point CR items.

The current study may be divided into two important phases:

1. A propensity score matching approach was conducted to make the CBT and PBT samples more comparable on all observed covariates that may affect student performance, aside from the test mode itself.
2. The comparability analyses were conducted on the matched samples, and conclusions were drawn based on the detected patterns.

Section R.2. Method

R.2.1. Preparing Balanced Samples

R.2.1.1. Overview

While the ideal conditions under which to investigate test mode comparability would necessitate random assignment of schools to test in either the CBT or PBT modes, the practical constraints and resources of individual districts and schools preclude such designs. The next best solution is often referred to as a quasi-experimental design. One popular such design involves propensity score matching (Austin, 2011a; Rosenbaum, 2010), which represents an additional step taken prior to studying test mode comparability. In other words, effective propensity score-matching procedures produce samples of PBT and CBT students that are, on average, otherwise comparable, with the only observed difference being that each sample tested in differing modes.

Table R.2.1 shows the number of students in the cleaned datasets by test mode prior to matching. This study relied on the same data-cleaning procedures that have been used for operational equating analyses, with the following additional rules:

- For Grades 4–8, students without scale scores in the same subject on the adjacent lower grade from the 2016 administration were removed.
- Because of sample size concerns and concerns about effects unrelated to test mode interfering with the study’s inferences, students testing with the Braille or large print forms were dropped, as well as were students who used a non-English language translation (e.g., Chinese, Korean, Haitian-Creole, Russian, Spanish) of a Mathematics form.

Table R.2.1. Sample Sizes Before Matching by Mode: ELA

Grade	Students	
	PBT	CBT
3	170,186	3,959
4	153,085	2,405
5	146,901	2,201
6	140,390	1,876
7	134,430	2,754
8	125,230	1,707

Note. Sample sizes indicate the number of students who took at least one item administered in both CBT and PBT, after the initial data cleaning used for OP equating and the additional data cleaning for the current study.

Table R.2.2. Sample Sizes Before Matching by Mode: Mathematics

Grade	Students	
	PBT	CBT
3	171,016	2,536
4	153,507	1,278
5	147,138	1,411
6	137,658	1,786
7	123,763	1,771
8	89,816	778

Note. Sample sizes indicate the number of students who took at least one item administered in both CBT and PBT, after the initial data cleaning used for OP equating and the additional data cleaning for the current study.

R.2.1.2. *Available Covariates*

The following covariates were used for estimating the propensity score model and, therefore, were directly balanced throughout the process of propensity score matching:

- grade $n - 1$ 2016 scale score;
- district-level number of CBT-eligible devices²;
- district-level ratio of CBT-eligible devices to enrolled students³;
- district-level minimum bandwidth entering a school⁴;
- student gender;
- student racial/ethnic category;
- student English language learner (ELL) status;
- student disability status;
- school-type (e.g., public, charter, religious and independent);
- district-level needs/resource capacity (NRC) code; and
- school-level region as specified by the joint management team definitions.

Questar also evaluated two-way interactions between the above covariates, but did not achieve any improvement in covariate balance in the final matched samples; therefore, simpler propensity scores models were employed. See **Appendix R.A: Propensity Score Models and Matching** for more detail on the propensity score models and additional matching results.

R.2.1.3. *Judging Covariate Balance*

The formulae for standardized differences in the context of propensity score matching are different for continuous and discrete variables, and there are minor modifications for estimating covariate balance before and after matching samples (Rosenbaum, 2010). The traditional experimental design is still a useful framework for this comparability study, so CBT can be considered the “treatment” and PBT can be considered the “control” condition. The analysis of covariate balance for discrete variables differs in that it uses the unbiased variance estimator for a proportion (see also page 174 of Austin (2011a) for examples of a similar, but not identical, formula).

For variable k being treated as **continuous**:

1. Estimate the means and variances for the treatment (\bar{x}_{tk} and s_{tk}^2) and control groups (\bar{x}_{ck} and s_{ck}^2) before matching.
2. Estimate the means only for the treatment (\bar{x}_{tmk}) and control groups (\bar{x}_{cmk}) after matching.
3. Estimate the standardized difference for variable k **before** matching as:

² This covariate was taken or derived from the 2016 New York State Education Department Instructional Technology Plan Survey. Since this survey is sent only to public school districts and for other reasons, there were between 50-60% missing data when evaluated across the entire 2017 operational equating sample.

³ See Footnote 2.

⁴ See Footnote 2.

$$d_{bk} = (\bar{x}_{tk} - \bar{x}_{ck}) / \sqrt{(s_{tk}^2 + s_{ck}^2)/2} \quad (1)$$

4. Estimate the standardized difference for variable k —note the use of the pre-matched pooled standard deviation in the denominator—**after** matching as:

$$d_{mk} = (\bar{x}_{tmk} - \bar{x}_{cmk}) / \sqrt{(s_{tk}^2 + s_{ck}^2)/2} \quad (2)$$

For variable k being treated as **discrete**:

1. Estimate the proportions for the treatment (p_{tk}) and control groups (p_{ck}) before matching.
2. Estimate the proportions for the treatment (p_{tmk}) and control groups (p_{cmk}) after matching.
3. Estimate the standardized difference for variable k **before** matching as:

$$d_{bk} = (p_{tk} - p_{ck}) / \sqrt{[p_{tk}(1 - p_{tk}) + p_{ck}(1 - p_{ck})]/2} \quad (3)$$

4. Estimate the standardized difference for variable k —note the use of the pre-matched pooled standard deviation in the denominator—**after** matching as:

$$d_{mk} = (p_{tmk} - p_{cmk}) / \sqrt{[p_{tk}(1 - p_{tk}) + p_{ck}(1 - p_{ck})]/2} \quad (4)$$

R.2.1.4. Propensity Score Models and Matching

In discussion with New York State’s Assessment Technical Advisory Committee (TAC), the decision was made to model the propensity score at the student level for CBT testing. The decision to adopt CBT was a school-level decision and modeling it at the student level violates one part of the assumption of strong ignorability (Rosenbaum and Rubin, 1983), meaning that some students had probabilities of assignment to CBT that equaled zero or one. By conditioning on student-level and school-level covariates, Questar was able to best approximate the selection process that one might observe if students were able to self-select and, therefore, treat school assignment as something that was ignorable.

The propensity score-matching process used a within-caliper matching approach, with caliper width defined as 0.02 times the standard deviation of the propensity score. This fine caliper was chosen because it did not cause a reduction in the number of matches made beyond the 0.2 level. The matching procedure was a one-to-one match without replacement (Austin 2011a).

R.2.2. Identifying Mode Effects

R.2.2.1. Evaluation of Test-level Mode Comparability

In order to detect test-level mode effects, Questar took the following two approaches after propensity score matching CBT and PBT students. First, the distribution of raw scores for the matched PBT and CBT samples were reviewed. This enabled a more direct means of detecting possible mode effects than comparing scale scores after equating.

When it comes to actually estimating the test mode effect, Questar used the single operational raw-score-to-scale-score (RSSS) conversion table (which was estimated based on all students in the operational equating sample), to predict scale scores for the matched CBT and matched PBT samples. The treatment effect was simply calculated as the difference in scale score means for the matched PBT and matched CBT samples.

Section R.3. Results

R.3.1. Propensity Score Matching

The purpose of applying propensity score matching was to achieve covariate balance on key factors that may influence test mode comparability. This study summarizes the covariate balance before and after matching, both graphically and in a tabular format. Standardized differences (d s) greater than 0.20 in absolute value will be flagged as not balanced, based on Cohen's (1992) labelling of $d = 0.20$ as a “small” effect. Appendix A presents further details on the propensity score model parameter estimates and the distribution of propensity scores.

R.3.1.1. *English Language Arts*

Prior to matching, there were a number of characteristics on the ELA samples that tended not to be balanced. Most notably, the proportion of New York City students testing on CBT [$M(d) = -1.183$]; the proportion of students whose district was missing responses for the 2016 New York State Education Department Instructional Technology Plan Survey [$M(d) = -1.074$]; and the proportion of students attending schools in average needs / resource districts [$M(d) = 0.892$]. After matching, the only characteristic that, on average, exceeds the 0.20 “small” effect size for standardized differences is the proportion of students enrolled in districts where the minimum bandwidth into a building is 10–49 Mbps [$M(d) = 0.262$], but this was a relatively small group size ($n = 56$ for the combined PBT and CBT matched samples).

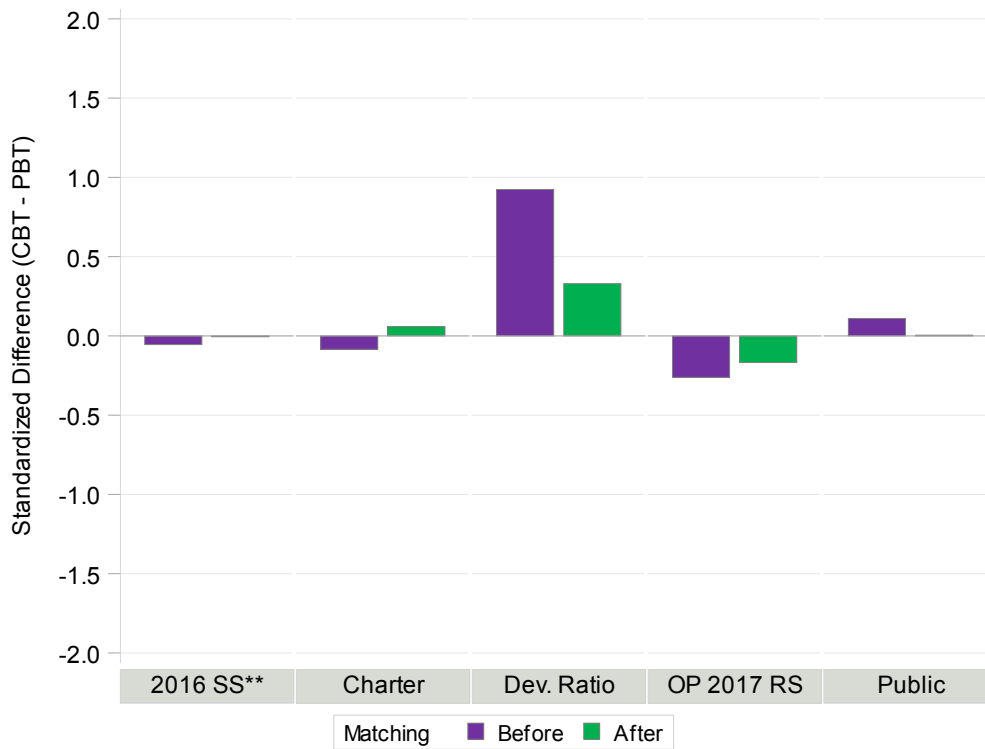
For ELA Grade 4, Table R.3.1 shows that there were two covariates for which, after propensity score matching, the standardized difference between PBT and CBT samples exceeded 0.20. Namely the district-level proportion of CBT-eligible devices to enrolled students ($d = 0.330$) and the proportion of students attending districts whose minimum bandwidth to a school building was greater than 10 Gbps ($d = 0.214$). The latter is associated with a rather small sample size of 89, but the former indicates that, even after matching, CBT students tended to attend districts with greater concentrations of devices that were eligible for CBT. In Figure R.3.1, key covariates were pulled out and presented graphically.

Table R.3.1. Covariate Balance Before and After Matching: ELA Grade 4

VariableValue		Before Matching			After Matching		
		n			n		
		PBT	CBT	d	PBT	CBT	d
Grade 3 2016 OP Scale Score		153,085	2,405	-0.055	2,405	2,405	-0.005
Total Eligible Devices in District		75,560	2,221	-0.225	2,218	2,221	0.010
District Devices / Enrollment		75,560	2,221	0.923	2,218	2,221	0.330
District Minimum Bandwidth	< 10 Mbps	771	10	-0.013	6	10	0.029
	10–49 Mbps	905	39	0.099	17	39	0.085
	50–99 Mbps	805	11	-0.010	23	11	-0.060
	100–999 Mbps	17,266	298	0.034	353	298	-0.067
	1–9 Gbps	41,859	1,530	0.782	1,515	1,530	0.013
	10 Gbps	12,034	254	0.093	294	254	-0.052
	> 10 Gbps	1,920	79	0.137	10	79	0.214
	Missing	77,525	184	-1.074	187	184	-0.005
Gender	Female	77,082	1,183	-0.023	1,164	1,183	0.016
	Male	76,003	1,222	0.023	1,241	1,222	-0.016
Ethnicity	Asian	16,724	101	-0.256	79	101	0.048
	Black	27,972	113	-0.436	108	113	0.010
	Hispanic	43,462	149	-0.614	128	149	0.037
	American Indian	1,015	22	0.028	125	107	-0.035
	Multiracial	3,842	85	0.060	0	0	n/r
	Pacific Islander	496	0	n/r	0	0	n/r
	White	59,574	1,935	0.935	1,965	1,935	-0.032
English Language Learner?	No	141,017	2,368	0.303	2,379	2,368	-0.040
	Yes	12,068	37	-0.303	26	37	0.040
Student with Disability?	No	133,531	2,160	0.081	2,184	2,160	-0.034
	Yes	19,554	245	-0.081	221	245	0.034
School Type	Public	136,505	2,221	0.110	2,218	2,221	0.005
	Charter	8,966	97	-0.084	71	97	0.059
	Religious and Independent	7,614	87	-0.067	116	87	-0.060
Needs/Resource Category (NRC) (only Public schools)	New York City	60,944	0	n/r	0	0	n/r
	Big 4 Cities	6,855	0	n/r	497	0	n/r
	Urban/Suburban	11,920	0	n/r	0	0	n/r
	High Needs Rural	8,150	493	0.465	0	493	n/r
	Average Needs	32,969	1,643	1.066	1,645	1,643	-0.002
	Low Needs	15,667	85	-0.267	76	85	0.021
High NRC?	No	65,216	1,912	0.817	1,908	1,912	0.004
	Yes	87,869	493	-0.817	497	493	-0.004
Joint Management Team Region	New York City	73,028	140	-1.074	156	140	-0.028
	Long Island	15,194	97	-0.233	97	97	0.000
	Lower Hudson Valley	12,538	0	n/r	0	0	n/r
	Mid-Hudson	6,656	158	0.098	160	158	-0.003
	Capital District / North Country	11,218	296	0.168	311	296	-0.019
	Central Region	3,644	0	n/r	0	0	n/r
	Mid-State	6,584	619	0.629	616	619	0.003
	Mid-South	4,764	237	0.276	208	237	0.042
	Mid-West	9,253	409	0.348	397	409	0.013
	West	10,182	449	0.367	460	449	-0.012
	Missing	24	0	n/r	0	0	n/r

Note. Bolded variables / groups had standardized differences (*d*) greater than 0.2 in absolute value. n/r: not reported due to sample size of fewer than five in one test mode sample.

Figure R.3.1. Key Covariate Balance Before and After Matching: ELA Grade 4



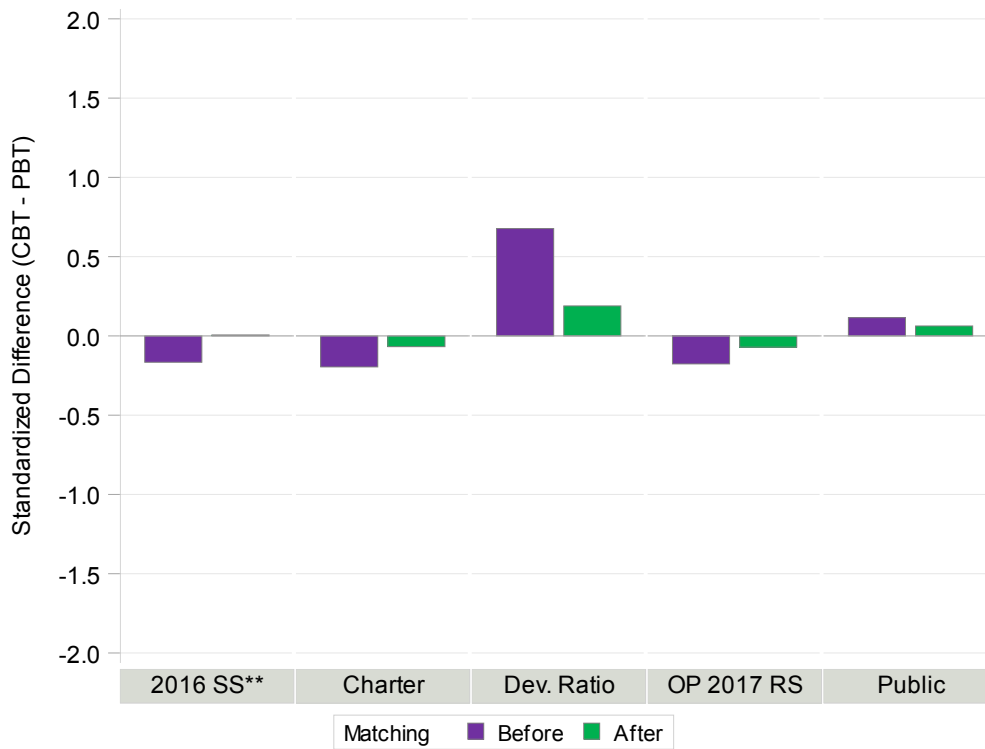
** Refers to Grade 3 2016 OP Scale Score.

For ELA Grade 5, Table R.3.2 shows that there were no covariates for which, after propensity score matching, the standardized difference between PBT and CBT samples exceeded 0.20. In Figure R.3.2, key covariates were pulled out and presented graphically.

Table R.3.2. Covariate Balance Before and After Matching: ELA Grade 5

Variable	Value	Before Matching			After Matching		
		n			n		
		PBT	CBT	d	PBT	CBT	d
Grade 4 2016 OP Scale Score		146,901	2,201	-0.166	2,201	2,201	0.006
Total Eligible Devices in District		71,289	2,028	-0.180	1,989	2,028	-0.062
District Devices / Enrollment		71,289	2,028	0.677	1,989	2,028	0.189
District Minimum Bandwidth	< 10 Mbps	740	7	-0.029	3	7	n/r
	10–49 Mbps	814	64	0.181	13	64	0.177
	50–99 Mbps	710	32	0.099	10	32	0.103
	100–999 Mbps	16,235	173	-0.109	260	173	-0.133
	1–9 Gbps	39,767	1,163	0.545	1,185	1,163	-0.020
	10 Gbps	11,183	569	0.504	434	569	0.147
	> 10 Gbps	1,840	20	-0.033	84	20	-0.192
	Missing	75,612	173	-1.087	212	173	-0.063
Gender	Female	73,365	1,064	-0.032	1,102	1,064	-0.035
	Male	73,536	1,137	0.032	1,099	1,137	0.035
Ethnicity	Asian	16,086	110	-0.221	90	115	0.054
	Black	27,701	159	-0.351	148	159	0.020
	Hispanic	41,383	169	-0.554	148	169	0.037
	American Indian	971	17	0.013	113	122	0.018
	Multiracial	3,179	105	0.143	0	0	n/r
	Pacific Islander	560	5	0.000	0	0	n/r
	White	57,021	1,636	0.767	1,702	1,636	-0.070
English Language Learner?	No	136,568	2,177	0.305	2,192	2,177	-0.079
	Yes	10,333	24	-0.305	9	24	0.079
Student with Disability?	No	127,025	1,984	0.114	1,967	1,984	0.025
	Yes	19,876	217	-0.114	234	217	-0.025
School Type	Public	130,391	2,028	0.115	1,989	2,028	0.063
	Charter	9,195	51	-0.196	76	51	-0.068
	Religious and Independent	7,315	122	0.025	136	122	-0.027
Needs/Resource Category (NRC) (only Public schools)	New York City	59,101	0	n/r	0	0	n/r
	Big 4 Cities	6,210	40	-0.141	520	629	0.113
	Urban/Suburban	11,060	168	0.004	0	0	n/r
	High Needs Rural	7,621	421	0.437	0	0	n/r
	Average Needs	31,159	1,388	0.936	1,458	1,388	-0.067
	Low Needs	15,240	11	-0.446	11	11	0.000
High NRC?	No	62,909	1,572	0.604	1,681	1,572	-0.113
	Yes	83,992	629	-0.604	520	629	0.113
Joint Management Team Region	New York City	71,214	82	-1.184	86	82	-0.009
	Long Island	14,457	213	-0.006	208	213	0.008
	Lower Hudson Valley	12,022	0	n/r	0	0	n/r
	Mid-Hudson	6,268	144	0.101	149	144	-0.009
	Capital District / North Country	10,694	248	0.138	250	248	-0.003
	Central Region	3,230	0	n/r	0	0	n/r
	Mid-State	5,921	832	0.913	787	832	0.042
	Mid-South	4,540	91	0.056	89	91	0.005
	Mid-West	8,850	379	0.355	401	379	-0.026
	West	9,684	212	0.112	231	212	-0.029
	Missing	21	0	n/r	0	0	n/r

Note. Bolded variables / groups had standardized differences (*d*) greater than 0.2 in absolute value. n/r: not reported due to sample size of fewer than five in one test mode sample.

Figure R.3.2. Key Covariate Balance Before and After Matching: ELA Grade 5

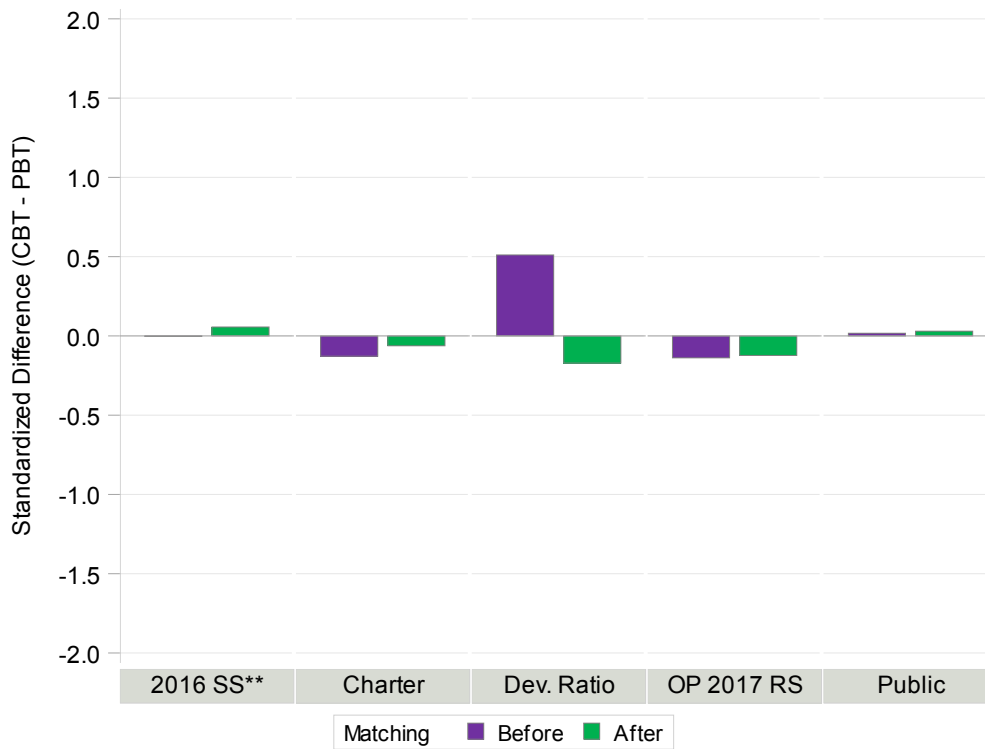
** Refers to Grade 4 2016 OP Scale Score.

For ELA Grade 6, Table R.3.3 shows that there was one covariate for which, after propensity score matching, the standardized difference between PBT and CBT samples exceeded 0.20: the proportion of students attending districts whose minimum bandwidth to a school building was between 10 and 49 Mbps ($d = 0.362$). Note that this standardized difference is associated with a rather small sample size of 129, so it should be interpreted with caution. In Figure R.3.3, key covariates were pulled out and presented graphically.

Table R.3.3. Covariate Balance Before and After Matching: ELA Grade 6

VariableValue		Before Matching			After Matching		
		n			n		
		PBT	CBT	d	PBT	CBT	d
Grade 5 2016 OP Scale Score		140,390	1,876	-0.003	1,683	1,683	0.056
Total Eligible Devices in District		65,958	1,662	-0.159	1,452	1,469	-0.062
District Devices / Enrollment		65,958	1,662	0.511	1,452	1,469	-0.171
District Minimum Bandwidth	< 10 Mbps	667	4	n/r	1	4	n/r
	10–49 Mbps	721	122	0.330	7	122	0.362
	50–99 Mbps	719	7	-0.021	13	7	-0.046
	100–999 Mbps	15,081	166	-0.064	236	166	-0.129
	1–9 Gbps	36,450	1,211	0.841	1,017	1,018	0.001
	10 Gbps	10,602	130	-0.024	140	130	-0.022
	> 10 Gbps	1,718	22	-0.005	38	22	-0.072
	Missing	74,432	214	-0.994	231	214	-0.030
Gender	Female	69,573	917	-0.014	794	815	0.025
	Male	70,817	959	0.014	889	868	-0.025
Ethnicity	Asian	15,661	86	-0.246	58	67	0.028
	Black	27,432	113	-0.413	100	113	0.032
	Hispanic	39,777	113	-0.619	83	113	0.076
	American Indian	955	18	0.031	48	59	0.037
	Multiracial	2,611	56	0.073	0	0	n/r
	Pacific Islander	414	0	n/r	0	0	n/r
	White	53,540	1,490	0.924	1,394	1,331	-0.095
English Language Learner?	No	131,468	1,848	0.252	1,656	1,656	0.000
	Yes	8,922	28	-0.252	27	27	0.000
Student with Disability?	No	119,994	1,679	0.122	1,496	1,502	0.011
	Yes	20,396	197	-0.122	187	181	-0.011
School Type	Public	123,620	1,662	0.017	1,452	1,469	0.030
	Charter	9,584	74	-0.128	97	74	-0.062
	Religious and Independent	7,186	140	0.097	134	140	0.013
Needs/Resource Category (NRC) (only Public schools)	New York City	57,661	0	n/r	0	0	n/r
	Big 4 Cities	5,755	49	-0.083	400	430	0.041
	Urban/Suburban	9,990	0	n/r	0	0	n/r
	High Needs Rural	6,854	381	0.478	0	0	n/r
	Average Needs	28,810	1,218	1.004	1,036	1,025	-0.013
	Low Needs	14,550	14	-0.429	16	14	-0.013
High NRC?	No	60,130	1,446	0.746	1,283	1,253	-0.041
	Yes	80,260	430	-0.746	400	430	0.041
Joint Management Team Region	New York City	70,143	156	-1.031	175	156	-0.038
	Long Island	13,483	10	-0.423	18	10	-0.052
	Lower Hudson Valley	11,373	0	n/r	0	0	n/r
	Mid-Hudson	5,717	122	0.109	86	122	0.089
	Capital District / North Country	9,803	364	0.373	358	364	0.009
	Central Region	2,990	0	n/r	0	0	n/r
	Mid-State	5,502	713	0.922	530	520	-0.013
	Mid-South	4,451	26	-0.120	34	26	-0.036
	Mid-West	7,959	276	0.302	263	276	0.021
	West	8,956	209	0.169	219	209	-0.018
	Missing	13	0	n/r	0	0	n/r

Note. Bolded variables / groups had standardized differences (*d*) greater than 0.2 in absolute value. n/r: not reported due to sample size of fewer than five in one test mode sample.

Figure R.3.3. Key Covariate Balance Before and After Matching: ELA Grade 6

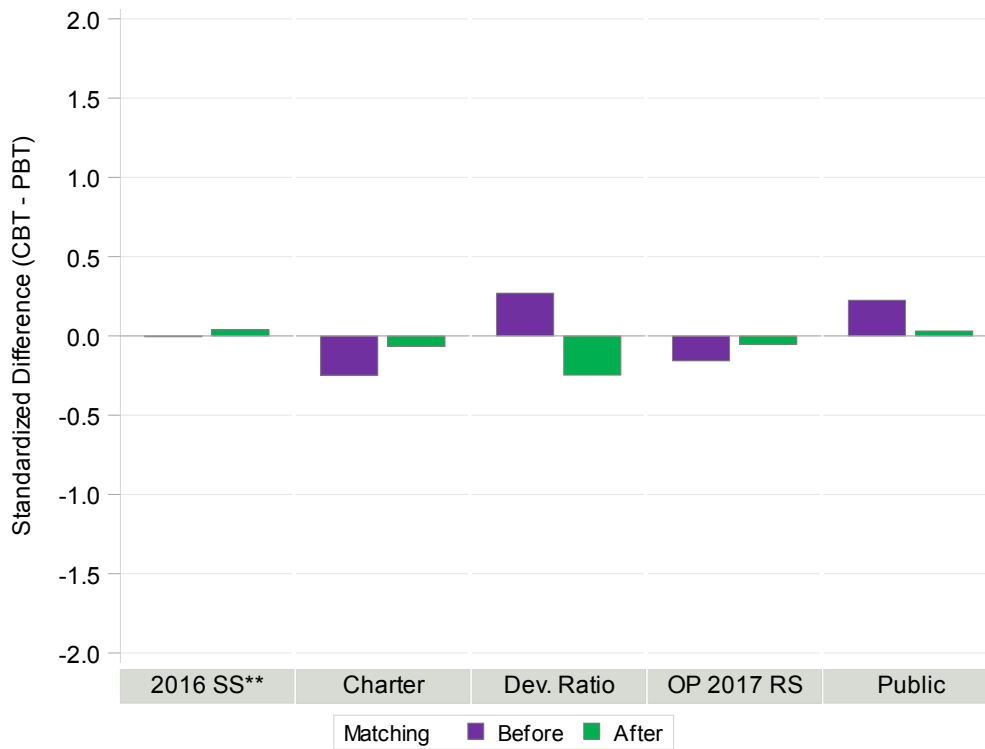
** Refers to Grade 5 2016 OP Scale Score.

For ELA Grade 7, Table R.3.4 shows that there were two covariates for which, after propensity score matching, the standardized difference between PBT and CBT samples exceeded 0.20. Namely the district-level proportion of CBT-eligible devices to enrolled students ($d = -0.247$) and the proportion of students attending districts whose minimum bandwidth to a school building was between 10 and 49 Mbps ($d = 0.300$). The latter is associated with a rather small sample size of 152, but the former indicates that, before matching, CBT students tended to attend districts with *greater* concentrations of devices that were eligible for CBT, after matching, there tended to be *lesser* concentrations of eligible devices in matched CBT students' districts. In Figure R.3.4, key covariates were pulled out and presented graphically.

Table R.3.4. Covariate Balance Before and After Matching: ELA Grade 7

Variable	Value	Before Matching			After Matching		
		n			n		
		PBT	CBT	d	PBT	CBT	d
Grade 6 2016 OP Scale Score		134,430	2,754	-0.006	2,754	2,754	0.038
Total Eligible Devices in District		60,570	2,587	-0.010	2,566	2,587	0.005
District Devices / Enrollment		60,570	2,587	0.270	2,566	2,587	-0.247
District Minimum Bandwidth	< 10 Mbps	668	0	n/r	30	0	n/r
	10–49 Mbps	631	143	0.288	9	143	0.300
	50–99 Mbps	649	24	0.047	5	24	0.000
	100–999 Mbps	13,601	352	0.084	451	352	-0.102
	1–9 Gbps	33,755	1,652	0.754	1,635	1,652	0.013
	10 Gbps	9,477	390	0.233	366	390	0.025
	> 10 Gbps	1,789	26	-0.036	70	26	-0.122
	Missing	73,860	167	-1.253	188	167	-0.031
Gender	Female	66,513	1,371	0.006	1,352	1,371	0.014
	Male	67,917	1,383	-0.006	1,402	1,383	-0.014
Ethnicity	Asian	15,789	145	-0.234	114	146	0.055
	Black	26,176	212	-0.349	212	212	0.000
	Hispanic	36,935	319	-0.409	274	319	0.053
	American Indian	958	12	-0.037	64	80	0.036
	Multiracial	2,099	68	0.065	0	0	n/r
	Pacific Islander	387	1	n/r	0	0	n/r
	White	52,086	1,997	0.723	2,090	1,997	-0.077
English Language Learner?	No	126,046	2,687	0.188	2,706	2,687	-0.048
	Yes	8,384	67	-0.188	48	67	0.048
Student with Disability?	No	115,070	2,496	0.156	2,509	2,496	-0.016
	Yes	19,360	258	-0.156	245	258	0.016
School Type	Public	117,547	2,587	0.225	2,566	2,587	0.031
	Charter	9,345	52	-0.248	80	52	-0.067
	Religious and Independent	7,538	115	-0.066	108	115	0.013
Needs/Resource Category (NRC) (only Public schools)	New York City	56,977	0	n/r	0	0	n/r
	Big 4 Cities	5,101	155	0.087	877	866	-0.009
	Urban/Suburban	8,427	344	0.215	0	0	n/r
	High Needs Rural	6,396	367	0.302	0	0	n/r
	Average Needs	25,707	1,478	0.769	1,421	1,478	0.041
	Low Needs	14,939	243	-0.076	268	243	-0.031
High NRC?	No	57,529	1,888	0.537	1,877	1,888	0.009
	Yes	76,901	866	-0.537	877	866	-0.009
Joint Management Team Region	New York City	69,972	120	-1.250	346	403	0.060
	Long Island	12,310	319	0.080	352	319	-0.037
	Lower Hudson Valley	10,582	283	0.084	0	0	n/r
	Mid-Hudson	5,300	47	-0.135	61	47	-0.037
	Capital District / North Country	9,427	213	0.028	225	213	-0.016
	Central Region	2,798	0	n/r	0	0	n/r
	Mid-State	4,993	870	0.786	892	870	-0.017
	Mid-South	3,915	254	0.267	200	254	0.071
	Mid-West	6,710	480	0.402	501	480	-0.020
	West	8,409	168	-0.006	177	168	-0.013
	Missing	14	0	n/r	0	0	n/r

Note. Bolded variables / groups had standardized differences (*d*) greater than 0.2 in absolute value. n/r: not reported due to sample size of fewer than five in one test mode sample.

Figure R.3.4. Key Covariate Balance Before and After Matching: ELA Grade 7

** Refers to Grade 6 2016 OP Scale Score.

For ELA Grade 8, Table R.3.5 shows that there was one covariate for which, after propensity score matching, the standardized difference between PBT and CBT samples exceeded 0.20: the proportion of students attending districts whose minimum bandwidth to a school building was between 10 and 49 Mbps ($d = 0.383$). The latter is associated with a rather small sample size of 146, so it should be interpreted with caution. In Figure R.3.5, key covariates were pulled out and presented graphically.

Table R.3.5. Covariate Balance Before and After Matching: ELA Grade 8

Variable	Value	Before Matching			After Matching		
		n			n		
		PBT	CBT	d	PBT	CBT	d
Grade 7 2016 OP Scale Score		125,230	1,707	-0.007	1,707	1,707	0.012
Total Eligible Devices in District		54,574	1,606	0.209	1,601	1,606	0.073
District Devices / Enrollment		54,574	1,606	0.666	1,601	1,606	0.111
District Minimum Bandwidth	< 10 Mbps	494	0	n/r	25	0	n/r
	10–49 Mbps	628	138	0.381	8	138	0.383
	50–99 Mbps	547	0	n/r	5	0	n/r
	100–999 Mbps	12,154	179	0.026	287	179	-0.185
	1–9 Gbps	30,848	957	0.676	939	957	0.021
	10 Gbps	8,241	304	0.348	290	304	0.022
	> 10 Gbps	1,662	28	0.026	47	28	-0.076
	Missing	70,656	101	-1.301	106	101	-0.012
Gender	Female	61,374	814	-0.026	783	814	0.036
	Male	63,856	893	0.026	924	893	-0.036
Ethnicity	Asian	15,039	140	-0.127	126	140	0.031
	Black	25,995	118	-0.409	109	118	0.021
	Hispanic	35,288	152	-0.512	129	152	0.049
	American Indian	968	2	n/r	31	38	0.029
	Multiracial	1,638	36	0.062	0	0	n/r
	Pacific Islander	380	0	n/r	0	0	n/r
	White	45,922	1,259	0.804	1,312	1,259	-0.072
English Language Learner?	No	118,199	1,654	0.123	1,656	1,654	-0.007
	Yes	7,031	53	-0.123	51	53	0.007
Student with Disability?	No	107,442	1,532	0.121	1,553	1,532	-0.042
	Yes	17,788	175	-0.121	154	175	0.042
School Type	Public	111,654	1,606	0.178	1,601	1,606	0.012
	Charter	7,835	0	n/r	32	0	n/r
	Religious and Independent	5,741	101	0.060	74	101	0.072
Needs/Resource Category (NRC) (only Public schools)	New York City	57,080	0	n/r	0	0	n/r
	Big 4 Cities	4,996	166	0.228	460	422	-0.051
	Urban/Suburban	7,979	0	n/r	0	0	n/r
	High Needs Rural	6,039	256	0.346	0	0	n/r
	Average Needs	23,070	839	0.687	819	839	0.023
	Low Needs	12,490	345	0.289	322	345	0.034
High NRC?	No	49,136	1,285	0.782	1,247	1,285	0.051
	Yes	76,094	422	-0.782	460	422	-0.051
Joint Management Team Region	New York City	67,162	46	-1.374	138	146	0.017
	Long Island	10,001	243	0.200	235	243	0.014
	Lower Hudson Valley	9,477	100	-0.068	0	0	n/r
	Mid-Hudson	4,751	100	0.096	102	100	-0.005
	Capital District / North Country	8,753	173	0.113	166	173	0.014
	Central Region	2,478	0	n/r	0	0	n/r
	Mid-State	4,830	500	0.728	494	500	0.008
	Mid-South	3,786	53	0.005	55	53	-0.007
	Mid-West	6,158	389	0.536	410	389	-0.029
	West	7,819	103	-0.009	107	103	-0.010
	Missing	15	0	n/r	0	0	n/r

Note. Bolded variables / groups had standardized differences (*d*) greater than 0.2 in absolute value. n/r: not reported due to sample size of fewer than five in one test mode sample.

Figure R.3.5. Key Covariate Balance Before and After Matching: ELA Grade 8

** Refers to Grade 7 2016 OP Scale Score.

R.3.1.2. *Mathematics*

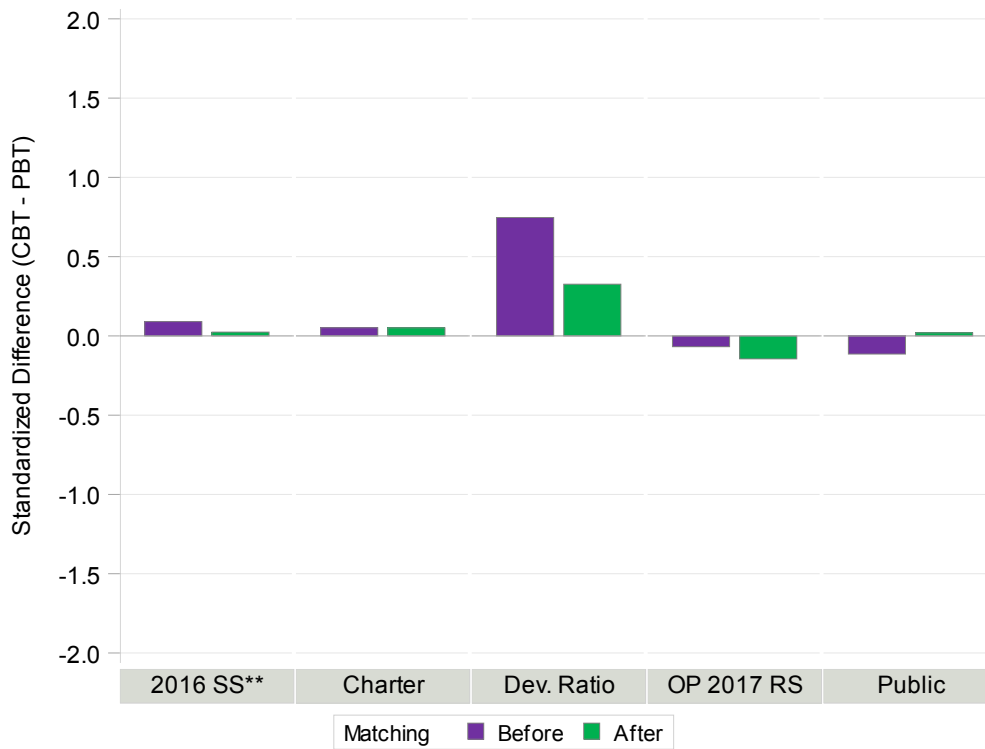
Prior to matching, there were a number of covariates on which the mathematics samples tended not to be balanced. Most notably, the proportion of New York City students using CBT [$M(d) = -1.149$]; the proportion of students whose district was missing responses for the 2016 New York State Education Department Instructional Technology Plan Survey [$M(d) = -1.081$]; and the proportion of students attending schools in average needs / resource districts [$M(d) = 1.002$]. After matching, no covariate, on average, exceeded the 0.20 “small” effect size for standardized differences; contrasting with ELA, the proportion of students enrolled in districts where the minimum bandwidth into a building is 10–49 Mbps did not pass the 0.20 threshold [$M(d) = 0.177$].

For mathematics Grade 4, Table R.3.6 shows that there was one covariate for which, after propensity score matching, the standardized difference between PBT and CBT samples exceeded 0.2. Namely the district-level proportion of CBT-eligible devices to enrolled students ($d = -0.326$). This indicates that after matching, CBT students tended to attend districts with greater concentrations of devices that were eligible for CBT. In Figure R.3.6, key covariates were pulled out and presented graphically.

Table R.3.6. Covariate Balance Before and After Matching: Mathematics Grade 4

Variable	Value	Before Matching			After Matching		
		n			n		
		PBT	CBT	d	PBT	CBT	d
Grade 3 2016 OP Scale Score		153,507	1,278	0.090	1,265	1,265	0.024
Total Eligible Devices in District		74,344	1,100	-0.279	1,079	1,087	0.047
District Devices / Enrollment		74,344	1,100	0.747	1,079	1,087	0.326
District Minimum Bandwidth	< 10 Mbps	748	9	0.028	9	9	0.000
	10–49 Mbps	878	36	0.175	6	36	0.184
	50–99 Mbps	787	10	0.034	11	10	-0.010
	100–999 Mbps	16,860	219	0.178	253	219	-0.078
	1–9 Gbps	41,314	729	0.641	690	716	0.044
	10 Gbps	11,838	74	-0.077	103	74	-0.091
	> 10 Gbps	1,919	23	0.045	7	23	0.103
	Missing	79,163	178	-0.876	186	178	-0.015
Gender	Female	76,061	595	-0.060	602	587	-0.024
	Male	77,446	683	0.060	663	678	0.024
Ethnicity	Asian	17,051	89	-0.145	69	76	0.019
	Black	28,500	77	-0.389	62	77	0.037
	Hispanic	43,245	97	-0.558	85	97	0.026
	American Indian	1,047	16	0.058	46	59	0.105
	Multiracial	3,790	43	0.053	0	0	n/r
	Pacific Islander	510	0	n/r	0	0	n/r
	White	59,364	956	0.783	1,003	956	-0.081
English Language Learner?	No	141,716	1,257	0.289	1,252	1,244	-0.030
	Yes	11,791	21	-0.289	13	21	0.030
Student with Disability?	No	131,427	1,179	0.213	1,180	1,166	-0.035
	Yes	22,080	99	-0.213	85	99	0.035
School Type	Public	137,836	1,100	-0.114	1,079	1,087	0.019
	Charter	9,201	93	0.052	77	93	0.051
	Religious and Independent	6,470	85	0.108	109	85	-0.084
Needs/Resource Category (NRC) (only Public schools)	New York City	63,491	0	n/r	0	0	n/r
	Big 4 Cities	6,656	0	n/r	215	0	n/r
	Urban/Suburban	11,112	0	n/r	0	0	n/r
	High Needs Rural	8,080	210	0.365	0	210	n/r
	Average Needs	32,914	800	0.917	775	787	0.021
	Low Needs	15,583	90	-0.111	89	90	0.003
High NRC?	No	64,168	1,068	0.957	1,050	1,055	0.009
	Yes	89,339	210	-0.957	215	210	-0.009
Joint Management Team Region	New York City	74,849	143	-0.899	146	143	-0.006
	Long Island	14,473	90	-0.087	84	90	0.017
	Lower Hudson Valley	12,255	0	n/r	0	0	n/r
	Mid-Hudson	6,196	153	0.296	160	153	-0.021
	Capital District / North Country	11,120	204	0.275	209	204	-0.012
	Central Region	3,544	0	n/r	0	0	n/r
	Mid-State	6,683	434	0.812	416	421	0.011
	Mid-South	4,845	43	0.012	41	43	0.009
	Mid-West	9,276	87	0.031	83	87	0.013
	West	10,240	124	0.111	126	124	-0.006
	Missing	26	0	n/r	0	0	n/r

Note. Bolded variables / groups had standardized differences (*d*) greater than 0.2 in absolute value. n/r: not reported due to sample size of fewer than five in one test mode sample.

Figure R.3.6. Key Covariate Balance Before and After Matching: Mathematics Grade 4

** Refers to Grade 3 2016 OP Scale Score.

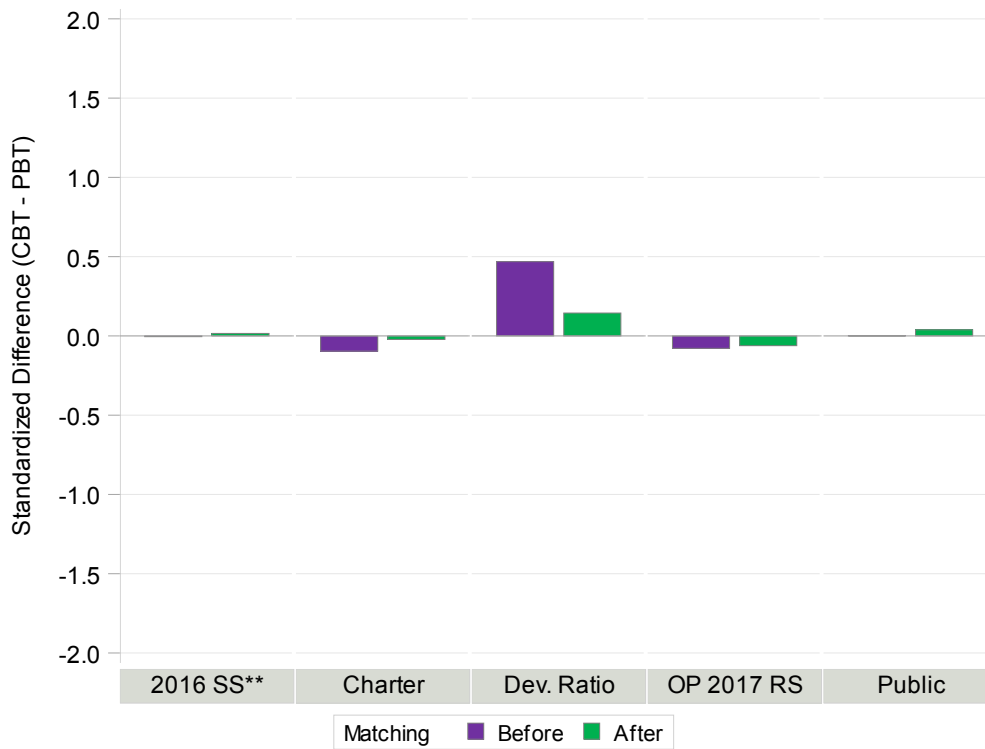
For mathematics Grade 5, Table R.3.7 shows that there was one covariate for which, after propensity score matching, the standardized difference between PBT and CBT samples exceeded 0.2. Namely the proportion of students attending districts whose minimum bandwidth to a school building was 100-999 Mbps ($d = -0.222$). Note that this standardized difference is associated with a relatively small sample size of 200 and should be interpreted with caution. In Figure R.3.7, key covariates were pulled out and presented graphically.

Table R.3.7. Covariate Balance Before and After Matching: Mathematics Grade 5

Variable	Value	Before Matching			After Matching		
		n			n		
		PBT	CBT	d	PBT	CBT	d
Grade 4 2016 OP Scale Score		147,138	1,411	-0.005	1,411	1,411	0.016
Total Eligible Devices in District		69,512	1,256	-0.398	1,239	1,256	-0.131
District Devices / Enrollment		69,512	1,256	0.468	1,239	1,256	0.144
District Minimum Bandwidth	< 10 Mbps	652	8	0.017	1	8	n/r
	10–49 Mbps	785	60	0.245	22	60	0.178
	50–99 Mbps	699	30	0.146	9	30	0.132
	100–999 Mbps	15,660	200	0.107	303	200	-0.222
	1–9 Gbps	38,813	669	0.447	670	669	-0.002
	10 Gbps	11,105	271	0.348	203	271	0.144
	> 10 Gbps	1,798	18	0.005	31	18	-0.083
	Missing	77,626	155	-1.003	172	155	-0.029
Gender	Female	72,559	683	-0.018	686	683	-0.004
	Male	74,579	728	0.018	725	728	0.004
Ethnicity	Asian	16,252	78	-0.201	94	80	-0.036
	Black	28,059	64	-0.462	52	64	0.027
	Hispanic	41,079	96	-0.580	79	96	0.033
	American Indian	988	14	0.035	82	96	0.109
	Multiracial	3,126	82	0.190	0	0	n/r
	Pacific Islander	582	2	n/r	0	0	n/r
	White	57,052	1,075	0.818	1,104	1,075	-0.045
English Language Learner?	No	137,090	1,398	0.310	1,405	1,398	-0.026
	Yes	10,048	13	-0.310	6	13	0.026
Student with Disability?	No	125,211	1,260	0.126	1,265	1,260	-0.011
	Yes	21,927	151	-0.126	146	151	0.011
School Type	Public	130,953	1,256	0.000	1,239	1,256	0.039
	Charter	9,298	59	-0.096	66	59	-0.022
	Religious and Independent	6,887	96	0.091	106	96	-0.030
Needs/Resource Category (NRC) (only Public schools)	New York City	61,441	0	n/r	0	0	n/r
	Big 4 Cities	6,041	0	n/r	382	0	n/r
	Urban/Suburban	10,185	154	0.140	0	448	n/r
	High Needs Rural	7,374	294	0.485	0	0	n/r
	Average Needs	30,780	742	0.695	793	742	-0.079
	Low Needs	15,132	66	-0.214	64	66	0.005
High NRC?	No	62,097	963	0.543	1,029	963	-0.097
	Yes	85,041	448	-0.543	382	448	0.097
Joint Management Team Region	New York City	73,538	90	-1.108	121	121	0.000
	Long Island	13,696	64	-0.189	66	64	-0.006
	Lower Hudson Valley	11,859	0	n/r	0	0	n/r
	Mid-Hudson	5,771	139	0.236	145	139	-0.017
	Capital District / North Country	10,478	211	0.252	229	211	-0.041
	Central Region	3,070	31	0.008	0	0	n/r
	Mid-State	6,184	461	0.789	423	461	0.075
	Mid-South	4,379	109	0.212	112	109	-0.010
	Mid-West	8,734	179	0.234	176	179	0.007
	West	9,405	127	0.098	139	127	-0.032
	Missing	24	0	n/r	0	0	n/r

Note. Bolded variables / groups had standardized differences (*d*) greater than 0.2 in absolute value. n/r: not reported due to sample size of fewer than five in one test mode sample.

Figure R.3.7. Key Covariate Balance Before and After Matching: Mathematics Grade 5



** Refers to Grade 4 2016 OP Scale Score.

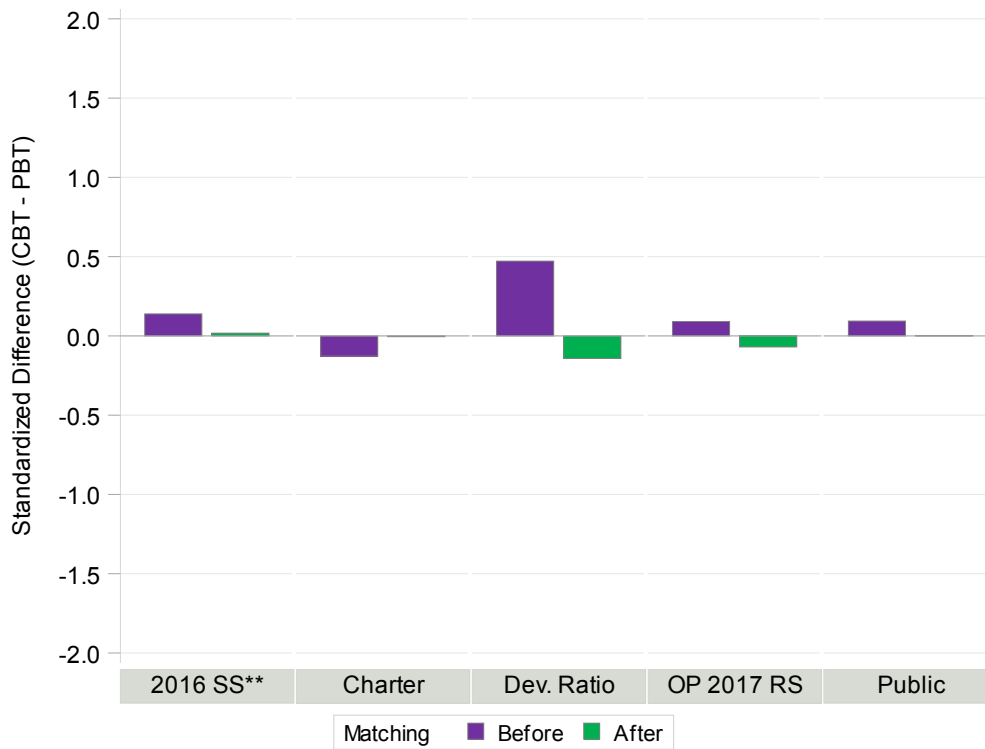
For mathematics Grade 6, Table R.3.8 shows that there were no covariates for which, after propensity score matching, the standardized difference between PBT and CBT samples exceeded 0.2. In Figure R.3.8, key covariates were pulled out and presented graphically.

Table R.3.8. Covariate Balance Before and After Matching: Mathematics Grade 6

VariableValue		Before Matching			After Matching		
		n			n		
		PBT	CBT	d	PBT	CBT	d
Grade 5 2016 OP Scale Score		137,658	1,786	0.139	1,762	1,762	0.017
Total Eligible Devices in District		62,523	1,610	-0.142	1,585	1,586	0.087
District Devices / Enrollment		62,523	1,610	0.472	1,585	1,586	-0.142
District Minimum Bandwidth	< 10 Mbps	565	29	0.121	0	29	n/r
	10–49 Mbps	763	43	0.154	9	43	0.160
	50–99 Mbps	684	7	-0.016	13	7	-0.051
	100–999 Mbps	14,202	183	-0.002	222	183	-0.073
	1–9 Gbps	34,503	1,203	0.937	1,180	1,179	-0.001
	10 Gbps	10,159	126	-0.013	132	126	-0.013
	> 10 Gbps	1,647	19	-0.013	29	19	-0.054
	Missing	75,135	176	-1.090	177	176	-0.001
Gender	Female	67,651	860	-0.020	842	847	0.006
	Male	70,007	926	0.020	920	915	-0.006
Ethnicity	Asian	15,530	84	-0.244	71	76	0.011
	Black	27,166	98	-0.439	77	98	0.037
	Hispanic	37,718	83	-0.652	83	83	0.000
	American Indian	973	11	-0.011	36	54	0.126
	Multiracial	2,500	53	0.075	0	0	n/r
	Pacific Islander	428	0	n/r	0	0	n/r
	White	53,343	1,457	0.973	1,495	1,451	-0.057
English Language Learner?	No	129,621	1,764	0.251	1,749	1,742	-0.022
	Yes	8,037	22	-0.251	13	20	0.022
Student with Disability?	No	117,277	1,633	0.195	1,603	1,611	0.014
	Yes	20,381	153	-0.195	159	151	-0.014
School Type	Public	120,006	1,610	0.094	1,585	1,586	0.002
	Charter	9,722	74	-0.127	76	74	-0.005
	Religious and Independent	7,930	102	-0.002	101	102	0.002
Needs/Resource Category (NRC) (only Public schools)	New York City	57,482	0	n/r	0	0	n/r
	Big 4 Cities	5,444	40	-0.099	272	298	0.085
	Urban/Suburban	8,880	0	n/r	0	0	n/r
	High Needs Rural	6,528	258	0.334	0	0	n/r
	Average Needs	27,357	1,275	1.208	1,274	1,251	-0.031
	Low Needs	14,315	37	-0.350	39	37	-0.005
High NRC?	No	59,324	1,488	0.918	1,490	1,464	-0.034
	Yes	78,334	298	-0.918	272	298	0.034
Joint Management Team Region	New York City	70,964	134	-1.103	167	162	-0.007
	Long Island	12,414	33	-0.320	34	33	-0.003
	Lower Hudson Valley	10,992	0	n/r	0	0	n/r
	Mid-Hudson	5,294	40	-0.094	47	40	-0.023
	Capital District / North Country	9,319	403	0.458	437	403	-0.056
	Central Region	2,845	28	-0.037	0	0	n/r
	Mid-State	5,430	658	0.894	582	634	0.080
	Mid-South	4,068	151	0.239	152	151	-0.002
	Mid-West	7,633	212	0.226	205	212	0.014
	West	8,686	127	0.032	138	127	-0.025
	Missing	13	0	n/r	0	0	n/r

Note. Bolded variables / groups had standardized differences (*d*) greater than 0.2 in absolute value. n/r: not reported due to sample size of fewer than five in one test mode sample.

Figure R.3.8. Key Covariate Balance Before and After Matching: Mathematics Grade 6



** Refers to Grade 5 2016 OP Scale Score.

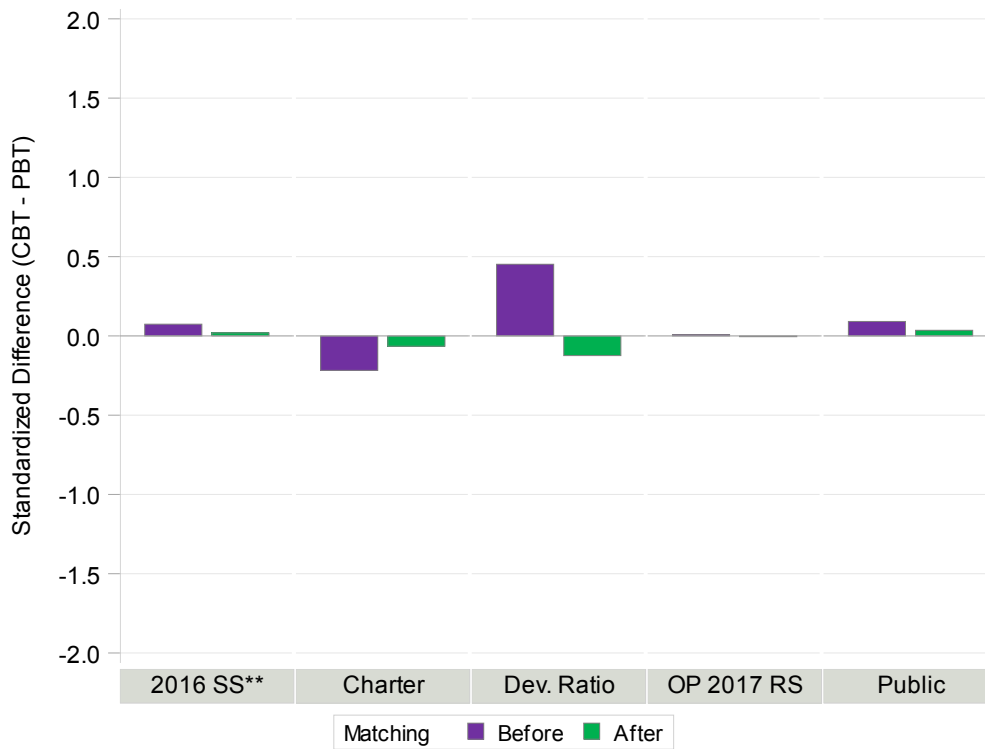
As with Grade 6, for mathematics Grade 7, Table R.3.9 shows that there were no covariates for which, after propensity score matching, the standardized difference between PBT and CBT samples exceeded 0.2. In Figure R.3.9, key covariates were pulled out and presented graphically.

Table R.3.9. Covariate Balance Before and After Matching: Mathematics Grade 7

Variable		Before Matching			After Matching		
		n			n		
		PBT	CBT	d	PBT	CBT	d
Grade 6 2016 OP Scale Score		123,763	1,771	0.074	1,748	1,748	0.019
Total Eligible Devices in District		52,641	1,609	-0.162	1,567	1,586	0.009
District Devices / Enrollment		52,641	1,609	0.452	1,567	1,586	-0.124
District Minimum Bandwidth	< 10 Mbps	584	0	n/r	5	0	n/r
	10–49 Mbps	633	39	0.147	6	39	0.164
	50–99 Mbps	616	21	0.075	11	21	0.063
	100–999 Mbps	9,872	256	0.206	243	256	0.024
	1–9 Gbps	30,278	1,064	0.773	1,063	1,041	-0.027
	10 Gbps	8,952	207	0.153	204	207	0.006
	> 10 Gbps	1,706	22	-0.012	35	22	-0.065
	Missing	71,122	162	-1.194	181	162	-0.027
Gender	Female	60,736	873	0.004	846	861	0.017
	Male	63,027	898	-0.004	902	887	-0.017
Ethnicity	Asian	15,148	110	-0.209	112	99	-0.026
	Black	25,061	121	-0.400	102	121	0.032
	Hispanic	33,746	188	-0.435	156	177	0.031
	American Indian	931	10	-0.023	37	60	0.163
	Multiracial	1,904	50	0.088	0	0	n/r
	Pacific Islander	366	1	n/r	0	0	n/r
	White	46,607	1,291	0.758	1,341	1,291	-0.062
English Language Learner?	No	117,322	1,757	0.261	1,739	1,734	-0.017
	Yes	6,441	14	-0.261	9	14	0.017
Student with Disability?	No	105,022	1,642	0.251	1,611	1,619	0.015
	Yes	18,741	129	-0.251	137	129	-0.015
School Type	Public	108,981	1,609	0.091	1,567	1,586	0.035
	Charter	9,339	49	-0.217	74	49	-0.065
	Religious and Independent	5,443	113	0.088	107	113	0.015
Needs/Resource Category (NRC) (only Public schools)	New York City	56,340	0	n/r	0	0	n/r
	Big 4 Cities	4,902	46	-0.077	577	553	-0.077
	Urban/Suburban	6,457	284	0.357	0	0	n/r
	High Needs Rural	5,762	225	0.289	0	0	n/r
	Average Needs	21,464	931	0.795	876	910	0.044
	Low Needs	14,056	123	-0.153	114	123	0.018
High NRC?	No	50,302	1,216	0.587	1,171	1,195	0.029
	Yes	73,461	555	-0.587	577	553	-0.029
Joint Management Team Region	New York City	67,583	116	-1.222	229	241	0.017
	Long Island	11,180	123	-0.077	122	123	0.002
	Lower Hudson Valley	10,195	125	-0.044	0	0	n/r
	Mid-Hudson	249	45	0.202	13	45	0.158
	Capital District / North Country	8,739	206	0.158	201	206	0.010
	Central Region	2,625	0	n/r	0	0	n/r
	Mid-State	4,913	643	0.881	650	620	-0.047
	Mid-South	3,776	79	0.074	84	79	-0.015
	Mid-West	6,494	295	0.372	290	295	0.009
	West	7,996	139	0.054	159	139	-0.044
	Missing	13	0	n/r	0	0	n/r

Note. Bolded variables / groups had standardized differences (*d*) greater than 0.2 in absolute value. n/r: not reported due to sample size of fewer than five in one test mode sample.

Figure R.3.9. Key Covariate Balance Before and After Matching: Mathematics Grade 7



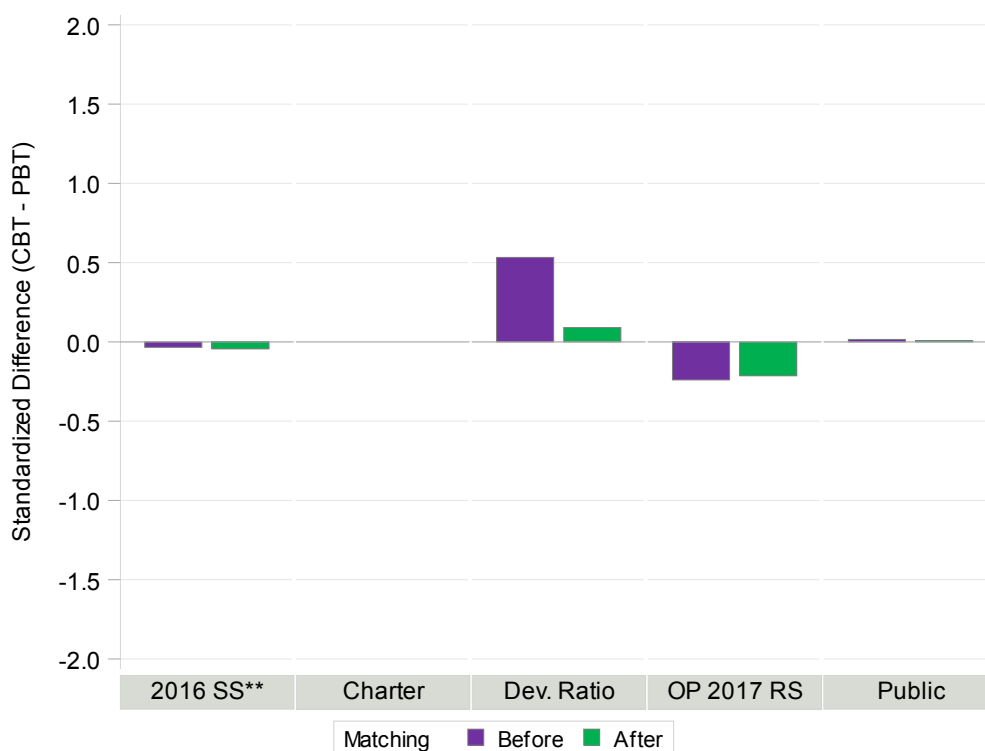
** Refers to Grade 6 2016 OP Scale Score.

For mathematics Grade 8, Table R.3.10 shows that there were no covariates for which, after propensity score matching, the standardized difference between PBT and CBT samples exceeded 0.2. In Figure R.3.10, key covariates were pulled out and presented graphically.

Table R.3.10. Covariate Balance Before and After Matching: Mathematics Grade 8

Variable	Value	Before Matching			After Matching		
		n		d	n		d
		PBT	CBT		PBT	CBT	
Grade 7 2016 OP Scale Score		89,816	778	-0.034	769	769	-0.044
Total Eligible Devices in District		32,255	679	-0.091	668	670	0.076
District Devices / Enrollment		32,255	679	0.533	668	670	0.090
District Minimum Bandwidth	< 10 Mbps	361	0	n/r	0	0	n/r
	10–49 Mbps	463	29	0.224	7	29	0.199
	50–99 Mbps	353	0	n/r	5	0	n/r
	100–999 Mbps	6,082	52	-0.004	55	52	-0.016
	1–9 Gbps	18,634	476	0.902	473	467	-0.017
	10 Gbps	5,253	106	0.265	114	106	-0.035
	> 10 Gbps	1,109	16	0.065	14	16	0.020
	Missing	57,561	99	-1.244	101	99	-0.006
Gender	Female	43,071	366	-0.018	368	365	-0.008
	Male	46,745	412	0.018	401	404	0.008
Ethnicity	Asian	9,267	65	-0.068	75	59	-0.072
	Black	20,765	43	-0.519	35	43	0.031
	Hispanic	27,437	41	-0.698	35	41	0.022
	American Indian	733	0	n/r	13	0	n/r
	Multiracial	1,114	21	0.105	0	18	n/r
	Pacific Islander	291	0	n/r	0	0	n/r
	White	30,209	608	1.003	611	608	-0.009
English Language Learner?	No	84,222	764	0.227	754	755	0.007
	Yes	5,594	14	-0.227	15	14	-0.007
Student with Disability?	No	74,211	667	0.085	664	659	-0.018
	Yes	15,605	111	-0.085	105	110	0.018
School Type	Public	77,919	679	0.016	668	670	0.008
	Charter	6,118	0	n/r	32	0	n/r
	Religious and Independent	5,779	99	0.215	69	99	0.133
Needs/Resource Category (NRC) (only Public schools)	New York City	45,664	0	n/r	0	0	n/r
	Big 4 Cities	4,184	37	0.005	130	128	-0.012
	Urban/Suburban	4,823	0	n/r	0	0	n/r
	High Needs Rural	4,330	91	0.252	0	0	n/r
	Average Needs	12,496	548	1.396	535	539	0.013
	Low Needs	6,422	3	n/r	3	3	n/r
High NRC?	No	30,815	650	1.156	639	641	0.006
	Yes	59,001	128	-1.156	130	128	-0.006
Joint Management Team Region	New York City	54,728	50	-1.412	53	50	-0.010
	Long Island	3,915	16	-0.131	16	16	0.000
	Lower Hudson Valley	7,247	0	n/r	0	0	n/r
	Mid-Hudson	270	16	0.163	0	16	n/r
	Capital District / North Country	6,026	158	0.406	159	158	-0.004
	Central Region	1,809	0	n/r	0	0	n/r
	Mid-State	3,305	279	0.883	264	270	0.021
	Mid-South	2,721	16	-0.062	17	16	-0.008
	Mid-West	3,934	164	0.518	164	164	0.000
	West	5,851	79	0.132	96	79	-0.080
	Missing	10	0	n/r	0	0	n/r

Note. Bolded variables / groups had standardized differences (*d*) greater than 0.2 in absolute value. n/r: not reported due to sample size of fewer than five in one test mode sample.

Figure R.3.10. Key Covariate Balance Before and After Matching: Mathematics Grade 8

** Refers to Grade 7 2016 OP Scale Score.

R.3.2. Test Mode Comparability Analyses

After having achieved reasonably good covariate balance between the matched CBT and PBT samples, the analysis of test mode comparability can proceed. We calculated the sample means for each matched sample and evaluated the differences and standardized differences for the following variables:

- Grade $n - 1$ 2016 Scale Score is the proxy for prior ability that was entered as a predictor into the propensity score model, and it is repeated in these analyses, in order to provide context.
- The 2017 operational raw score was evaluated overall and separately including only MC or only CR item scored responses.
- The 2017 scale score referenced in these analyses refers to that which comes from a single raw-score-to-scale-score conversion table.
 - These differences will serve as the basis for calculating a scale score adjustment to CBT students' scores, but at this point, only the single common table is used.
- The proportions of students falling into each performance level from I to IV are also presented.

Table R.3.11. Test Performance After Matching: ELA

Grade	Variable	Value	After Matching							
			PBT			CBT			Δ	d
			n	M	SD	n	M	SD		
4	Grade 3 2016 Scale Score		2,405	309.9	31.8	2,405	309.7	31.7	-0.160	-0.005
	2017 MC Raw Score		2,405	14.50	4.85	2,405	14.11	4.70	-0.384	-0.079
	2017 CR Raw Score		2,405	10.93	4.46	2,405	9.88	4.54	-1.043	-0.224
	2017 Raw Score		2,405	25.42	8.55	2,405	24.00	8.49	-1.427	-0.163
	2017 Scale Score		2,405	305.7	32.1	2,405	300.4	32.1	-5.269	-0.159
	2017 OP	NYS I	551	0.229		697	0.290		0.061	0.049
	Performance	NYS II	907	0.377		915	0.380		0.003	0.002
	Level	NYS III	649	0.270		579	0.241		-0.029	-0.024
		NYS IV	298	0.124		214	0.089		-0.035	-0.037
	Grade 4 2016 Scale Score		2,201	303.1	32.1	2,201	303.3	31.8	0.199	0.006
5	2017 MC Raw Score		2,201	20.71	6.29	2,201	20.61	6.01	-0.102	-0.017
	2017 CR Raw Score		2,201	11.26	4.62	2,201	10.66	4.50	-0.601	-0.129
	2017 Raw Score		2,201	31.97	10.13	2,201	31.27	9.76	-0.703	-0.070
	2017 Scale Score		2,201	300.5	37.0	2,201	298.1	35.2	-2.409	-0.066
	2017 OP	NYS I	727	0.330		782	0.355		0.025	0.019
	Performance	NYS II	729	0.331		767	0.348		0.017	0.013
	Level	NYS III	514	0.234		462	0.210		-0.024	-0.020
		NYS IV	231	0.105		190	0.086		-0.019	-0.021
	Grade 5 2016 Scale Score		1,683	296.2	36.7	1,683	298.3	37.9	2.099	0.055
	2017 MC Raw Score		1,683	22.08	6.86	1,683	21.68	7.04	-0.400	-0.056
6	2017 CR Raw Score		1,683	13.43	4.93	1,683	12.49	4.85	-0.939	-0.191
	2017 Raw Score		1,683	35.50	10.77	1,683	34.16	10.94	-1.339	-0.121
	2017 Scale Score		1,683	300.1	32.9	1,683	295.6	33.2	-4.562	-0.134
	2017 OP	NYS I	459	0.273		519	0.308		0.036	0.028
	Performance	NYS II	716	0.425		721	0.428		0.003	0.002
	Level	NYS III	262	0.156		272	0.162		0.006	0.006
		NYS IV	246	0.146		171	0.102		-0.045	-0.045
	Grade 6 2016 Scale Score		2,754	300.5	33.4	2,754	301.8	34.3	1.299	0.037
	2017 MC Raw Score		2,754	21.42	6.64	2,754	21.20	6.58	-0.228	-0.034
	2017 CR Raw Score		2,754	14.17	4.83	2,754	13.84	4.93	-0.331	-0.068
7	2017 Raw Score		2,754	35.59	10.50	2,754	35.04	10.65	-0.559	-0.052
	2017 Scale Score		2,754	307.2	31.2	2,754	305.5	31.6	-1.716	-0.054
	2017 OP	NYS I	582	0.211		634	0.230		0.019	0.016
	Performance	NYS II	1,071	0.389		1,060	0.385		-0.004	-0.003
	Level	NYS III	811	0.294		783	0.284		-0.010	-0.008
		NYS IV	290	0.105		277	0.101		-0.005	-0.005
	Grade 7 2016 Scale Score		1,707	304.2	34.5	1,707	304.6	33.5	0.413	0.012
	2017 MC Raw Score		1,707	22.90	6.98	1,707	22.69	6.83	-0.211	-0.031
	2017 CR Raw Score		1,707	15.38	4.71	1,707	15.52	4.65	0.136	0.030
	2017 Raw Score		1,707	38.28	10.78	1,707	38.20	10.61	-0.076	-0.007
8	2017 Scale Score		1,707	306.1	34.4	1,707	306.1	33.6	-0.079	-0.002
	2017 OP	NYS I	362	0.212		349	0.204		-0.008	-0.007
	Performance	NYS II	548	0.321		593	0.347		0.026	0.020
	Level	NYS III	573	0.336		523	0.306		-0.029	-0.022
		NYS IV	224	0.131		242	0.142		0.011	0.010
	Grade 8 2016 Scale Score		1,707	304.2	34.5	1,707	304.6	33.5	0.413	0.012
	2017 MC Raw Score		1,707	22.90	6.98	1,707	22.69	6.83	-0.211	-0.031
	2017 CR Raw Score		1,707	15.38	4.71	1,707	15.52	4.65	0.136	0.030
	2017 Raw Score		1,707	38.28	10.78	1,707	38.20	10.61	-0.076	-0.007
	2017 Scale Score		1,707	306.1	34.4	1,707	306.1	33.6	-0.079	-0.002

Note. Bolded variables / groups had standardized differences (d) greater than 0.1 in absolute value. Black shaded cells show the scale score adjustment factors before rounding.

Table R.3.12. Test Performance After Matching: Mathematics

Grade	Variable	Value	After Matching							
			PBT			CBT			Δ	d
			n	M	SD	n	M	SD		
4	Grade 3 2016 Scale Score		1,265	311.3	34.8	1,265	312.2	34.6	0.832	0.024
	2017 MC Raw Score		1,265	26.28	7.42	1,265	25.55	7.55	-0.728	-0.092
	2017 CR Raw Score		1,265	13.42	6.42	1,265	12.12	6.29	-1.299	-0.196
	2017 Raw Score		1,265	39.71	13.29	1,265	37.68	13.27	-2.027	-0.145
	2017 Scale Score		1,265	309.1	35.2	1,265	304.1	34.5	-5.026	-0.133
	2017 OP	NYS I	242	0.191		305	0.241		0.050	0.116
	Performance	NYS II	435	0.344		460	0.364		0.020	0.042
	Level	NYS III	326	0.258		305	0.241		-0.017	-0.039
		NYS IV	262	0.207		195	0.154		-0.053	-0.136
5	Grade 4 2016 Scale Score		1,411	306.9	39.0	1,411	307.5	36.3	0.590	0.016
	2017 MC Raw Score		1,411	24.89	8.22	1,411	24.69	8.06	-0.200	-0.024
	2017 CR Raw Score		1,411	9.65	6.61	1,411	8.96	6.30	-0.684	-0.102
	2017 Raw Score		1,411	34.54	14.27	1,411	33.65	13.68	-0.884	-0.062
	2017 Scale Score		1,411	309.7	35.6	1,411	307.6	33.9	-2.134	-0.059
	2017 OP	NYS I	393	0.279		406	0.288		0.009	0.020
	Performance	NYS II	392	0.278		426	0.302		0.024	0.054
	Level	NYS III	430	0.305		424	0.300		-0.004	-0.009
		NYS IV	196	0.139		155	0.110		-0.029	-0.084
6	Grade 5 2016 Scale Score		1,762	313.5	33.6	1,762	314.1	34.5	0.595	0.017
	2017 MC Raw Score		1,762	26.20	8.75	1,762	25.43	8.66	-0.771	-0.084
	2017 CR Raw Score		1,762	12.00	6.07	1,762	11.72	5.97	-0.274	-0.043
	2017 Raw Score		1,762	38.19	14.31	1,762	37.15	14.09	-1.044	-0.070
	2017 Scale Score		1,762	312.5	36.3	1,762	309.7	36.0	-2.817	-0.071
	2017 OP	NYS I	329	0.187		351	0.199		0.012	0.030
	Performance	NYS II	598	0.339		615	0.349		0.010	0.021
	Level	NYS III	418	0.237		440	0.250		0.012	0.030
		NYS IV	417	0.237		356	0.202		-0.035	-0.084
7	Grade 6 2016 Scale Score		1,748	310.0	38.0	1,748	310.7	35.8	0.712	0.019
	2017 MC Raw Score		1,748	23.71	9.33	1,748	23.74	9.35	0.026	0.003
	2017 CR Raw Score		1,748	9.98	6.95	1,748	9.88	6.84	-0.106	-0.015
	2017 Raw Score		1,748	33.69	15.71	1,748	33.61	15.59	-0.080	-0.005
	2017 Scale Score		1,748	311.1	32.3	1,748	310.8	32.1	-0.256	-0.007
	2017 OP	NYS I	417	0.239		418	0.239		0.001	0.001
	Performance	NYS II	611	0.350		602	0.344		-0.005	-0.011
	Level	NYS III	537	0.307		547	0.313		0.006	0.013
		NYS IV	183	0.105		181	0.104		-0.001	-0.003
8	Grade 7 2016 Scale Score		769	300.1	33.0	769	298.7	33.7	-1.475	-0.044
	2017 MC Raw Score		769	21.80	8.33	769	20.24	8.28	-1.564	-0.177
	2017 CR Raw Score		769	7.26	5.81	769	5.79	5.35	-1.473	-0.246
	2017 Raw Score		769	29.07	13.53	769	26.03	13.00	-3.038	-0.214
	2017 Scale Score		769	294.3	36.0	769	286.0	37.7	-8.270	-0.213
	2017 OP	NYS I	261	0.339		332	0.432		0.092	0.189
	Performance	NYS II	346	0.450		315	0.410		-0.040	-0.082
	Level	NYS III	137	0.178		106	0.138		-0.040	-0.111
		NYS IV	25	0.033		16	0.021		-0.012	-0.056

Note. Bolded variables / groups had standardized differences (d) greater than 0.1 in absolute value. Black shaded cells show the scale score adjustment factors before rounding.

R.3.2.1. *English Language Arts*

Table R.3.11 shows that after matching, the CBT and PBT students' mean scale scores in the prior grade's 2016 operational scale score metric was, at most, 2.099 (Grade 6) in absolute value, indicating that the samples were comparable on this measure of prior ability. The only outcome measure on which the matched CBT and PBT students differed (i.e., exceeded 0.20 in terms of absolute value of the standardized difference) was the CR raw score for Grade 4 ($d = -0.224$).

R.3.2.2. *Mathematics*

Table R.3.12 shows that after matching, the CBT and PBT students' mean scale scores in the prior grade's 2016 operational scale score metric was, at most, 1.475 (Grade 8) in absolute value, indicating that the samples were comparable on this measure of prior ability. The only grade for which the matched CBT and PBT students differed in terms of whether covariates exceed 0.20 in terms of absolute value of the standardized difference was Grade 8. In that grade, the 769 matched pairs tended to favor the matched PBT sample over the matched CBT sample in terms of: (a) CR raw score ($d = -0.246$); (b) overall raw score ($d = -0.214$); and (c) common scale score ($d = -0.213$).

Grade 8 mathematics is unlike the other 11 subject and grade combinations in that some advanced eighth grade students are exempted from taking the Grade 8 test, as a result of their taking the high school Regents examination in Algebra I. Beginning in 2014, the U.S. Department of Education granted NYSED a waiver permitting individual schools to decide whether their advanced eighth grade students taking a high school mathematics course would be required to also take the Grade 8 test. As such, the size of the Grade 8 mathematics test-taking population—and correspondingly the equating sample used for all analysis, including this study of test mode comparability—tends to have around 40,000 fewer students. Many of those students whose schools waive the requirement to take the Grade 8 test would have been generally more proficient in mathematics, so the change in sample is not just in the sheer size, but also in the proficiency distribution of the sample. For more information on the waiver, see: <http://www.nysed.gov/news/2015/new-york-granted-federal-waiver-eliminate-double-testing-math>.

Section R.4. Discussion and Conclusions

R.4.1. Discussion

Based on the analyses described above, NYSED—in consultation with New York State’s Assessment TAC and Questar—decided to apply an additive adjustment to CBT students’ scale scores. A number of options were considered, but an additive adjustment to scale scores was selected, as it best balanced concerns about fairness and interpretability / face validity. NYSED also chose to set a ceiling, above which the CBT students’ scale scores would not be adjusted: namely, the maximum observed scale score available to PBT students. In other words, the highest scale score on CBT was constrained to be equal to the highest scale score for PBT students.

Table R.4.1. CBT Scale Score Adjustments

Subject	Grade	After Matching							
		PBT			CBT			CBT	
		<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	Δ	<i>d</i> Adjustment
ELA	4	2,405	305.7	32.1	2,405	300.4	32.1	-5.269	-0.159 +5
	5	2,201	300.5	37.0	2,201	298.1	35.2	-2.409	-0.066 +2
	6	1,683	300.1	32.9	1,683	295.6	33.2	-4.562	-0.134 +5
	7	2,754	307.2	31.2	2,754	305.5	31.6	-1.716	-0.054 +2
	8	1,707	306.1	34.4	1,707	306.1	33.6	-0.079	-0.002 0
Math	4	1,265	309.1	35.2	1,265	304.1	34.5	-5.026	-0.133 +5
	5	1,411	309.7	35.6	1,411	307.6	33.9	-2.134	-0.059 +2
	6	1,762	312.5	36.3	1,762	309.7	36.0	-2.817	-0.071 +3
	7	1,748	311.1	32.3	1,748	310.8	32.1	-0.256	-0.007 0
	8	769	294.3	36.0	769	286.0	37.7	-8.270	-0.213 +8

Note. CBT scale scores were only adjusted up to the maximum observed PBT scale score value.

R.4.2. Conclusions

Only a small proportion of schools (corresponding to about 1.8% of students) chose to administer the Grades 3–8 New York Common Core ELA and Mathematics Tests via CBT. Therefore, the population of students who tested via CBT were not assumed equivalent to the population of students who tested via PBT. In order to select a sample of PBT students that could be compared to the population of CBT students, propensity score matching was conducted. The propensity score-matched results revealed small but meaningful differences in CBT performance, even when compared to a comparable group of PBT students.

Section R.5. References

- Austin, P. C. (2011a). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3), 399–424. doi: <http://dx.doi.org/10.1080/00273171.2011.568786>
- Austin, P. C. (2011b). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics*, 10(2), 150–161. doi: <http://dx.doi.org/10.1002/pst.433>
- Cai, L. Thissen, D., & du Toit, S. (2017). *IRTPRO 4* [Computer software]. Skokie, IL: Scientific Software International, Inc.
- Cohen, J. (1992). A power primer. *Psychological bulletin*, 112(1), 155. <http://dx.doi.org/10.1037/0033-2909.112.1.155>
- Firth, D. (1993). Bias reduction of maximum likelihood estimates, *Biometrika*, 80, 27–38. <http://dx.doi.org/10.1093/biomet/80.1.27>
- Kim, S. & M. J. Kolen (2004). *STUIRT: A computer program for scale transformation under unidimensional item response theory models* [Computer software]. Iowa City, IA: Iowa Testing Programs, The University of Iowa.
- Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Oxford, England: Addison-Wesley
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied psychological measurement*, 16(2), 159–176. <http://dx.doi.org/10.1177/014662169201600206>
- Rosenbaum, P. R. (2010). *Design of observational studies*. New York: Springer. doi: <http://dx.doi.org/10.1007/978-1-4419-1213-8>
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55. doi: <http://dx.doi.org/10.1093/biomet/70.1.41>
- Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7(2), 201–210. doi: <http://dx.doi.org/10.1177/014662168300700208xxxx>
- Thissen, D., Chen, W-H., & Bock, R. D. (2003). *MULTILOG 7 for Windows: Multiple-category item analysis and test scoring using item response theory* [Computer software]. Skokie, IL: Scientific Software International, Inc.
- Way, W. D., Davis, L. L., & Strain-Seymour, E. (2008). *The validity case for assessing direct writing by computer (A Pearson Assessments & Information White Paper)*. Pearson: Iowa City, IA.

Appendix R.A: Propensity Score Models and Matching

R.A.1. Technical Concerns for Propensity Score Modeling with Low CBT Adoption

In order to address the technical concerns with modeling a relatively low CBT adoption rate, two measures were taken. First, a penalized likelihood approach (Firth, 1993) was used for model estimation, in order to reduce the bias associated with covariates that are perfectly or almost perfectly related to CBT adoption, which would cause complete or quasi-complete separation, respectively. Second, where CBT adoption was zero or near zero, (e.g., for the New York City joint management team or JMT region code), certain levels of the covariates were combined, rather than dropping cases.

R.A.2. Propensity Score Model and Matching Results

R.A.2.1. Propensity Score Model Results

The propensity score models for each subject and grade are shown as Table R.A.1 through Table R.A.10. Some of the key findings include:

- Students' prior year scale score from the 2016 Operational administration in the same subject, but the lower adjacent grade.
- The two covariates derived from the 2016 New York State Education Department Instructional Technology Plan Survey items administered to district level staff. They provided meaningful insight, despite between 50–60% rates of missing data for the test taking populations.
 - The “District Minimum Bandwidth” covariate was a categorical survey item asking respondents to provide the minimum capacity “of the telecommunications line coming into the district’s school building(s) from the district hub or district data center.”
 - The “Ratio of CBT Devices to District Enrollment” variable was constructed by dividing the total devices reported by the district that could have been eligible for the 2017 Operational CBT administration by the total district enrollment.

R.A.2.2. Propensity Score Matching Results

In addition, the distributions of propensity scores for online- and paper-based testing schools both before and after matching are shown in Figure R.A.1 for ELA and Figure R.A.2 for mathematics.

R.A.3. Distribution of Raw Score Differences After Matching

Finally, when performing propensity score analyses, it is important to evaluate the extent to which the results that one produces are not idiosyncratic of the particular random seed selected for randomizing the CBT students for matching. As such, 100 replicates of the propensity score matching procedure were conducted, and the key outcome evaluated was the distribution of raw score differences between matched PBT and matched CBT students. These distributions are presented as Figure R.A.3 for ELA and Figure R.A.4 for mathematics. Note that the analyses in the body of this report are based on what is labeled as “Replicate 0.” Furthermore, the random seeds were themselves selected using a pseudo-random number generation process and the seeds identified for reporting were designated prior to the review of any results.

Table R.A.1. Propensity Score Model: ELA Grade 4

Variable	Value	Est. (SE)
Intercept		-4.276 (0.148)
Grade 3 2016 OP Scale Score*		-0.071 (0.026)
Ethnicity	Asian/Pacific Islander	-0.111 (0.109)
	Black	-1.329 (0.102)
	Hispanic	-0.869 (0.091)
	Other / Missing	-0.286 (0.103)
English Language Learner		-0.744 (0.174)
Student with Disability		-0.198 (0.079)
Needs/ Resource Category	High Needs	-0.889 (0.056)
	Low Needs	-1.831 (0.118)
	Charter / Religious and Independent	5.670 (1.415)
Joint Management Team Region	Capital District / North Country	-0.138 (0.092)
	Long Island	-0.711 (0.132)
	Mid-Hudson	0.053 (0.111)
	Mid-State	1.038 (0.082)
	Mid-West	0.143 (0.088)
	West	0.087 (0.084)
	Other / Missing	-1.325 (0.120)
District Minimum Bandwidth	1–9 Gbps	0.718 (0.064)
	>= 10 Gbps	0.185 (0.082)
	Missing	-4.567 (1.421)
Ratio of CBT Devices to District Enrollment	0.33 - 0.66	-0.388 (0.142)
	>= 0.67	1.408 (0.123)
	Missing	n/a

Note. *: Grand-mean centered. Bolded parameter estimates were significant at the $\alpha = .01$ level. Reference students were non-ELL and non-disabled White students attending average needs / resource public schools in the Mid-South JMT region. Their districts reported that the minimum building bandwidth was less than 1 Gbps and a ratio of CBT-eligible devices to enrolled students of less than 1/3. AIC: 24,705.3 (intercept only) and 18,492.1 (intercept and covariates).

Table R.A.2. Propensity Score Model: ELA Grade 5

Variable	Value	Est. (SE)
Intercept		-10.648 (1.417)
Grade 4 2016 OP Scale Score*		-0.134 (0.027)
Ethnicity	Asian/Pacific Islander	0.029 (0.104)
	Black	-0.985 (0.090)
	Hispanic	-0.822 (0.090)
	Other / Missing	0.096 (0.099)
English Language Learner		-1.334 (0.210)
Student with Disability		-0.361 (0.084)
Needs/ Resource	High Needs	-0.246 (0.054)
	Low Needs	-4.059 (0.299)
Category	Charter / Religious and Independent	5.889 (1.416)
Joint Management Team Region	Capital District / North Country	0.672 (0.125)
	Long Island	1.311 (0.133)
	Mid-Hudson	1.170 (0.139)
	Mid-State	2.321 (0.114)
	Mid-West	0.967 (0.121)
	West	0.247 (0.127)
	Other / Missing	-0.922 (0.164)
District Minimum Bandwidth	1–9 Gbps	1.038 (0.071)
	>= 10 Gbps	1.336 (0.079)
	Missing	0.745 (1.999)
Ratio of CBT	0.33 - 0.66	5.382 (1.412)
Devices to District	>= 0.67	6.109 (1.412)
Enrollment	Missing	n/a

Note. *: Grand-mean centered. Bolded parameter estimates were significant at the $\alpha = .01$ level. Reference students were non-ELL and non-disabled White students attending average needs / resource public schools in the Mid-South JMT region. Their districts reported that the minimum building bandwidth was less than 1 Gbps and a ratio of CBT-eligible devices to enrolled students of less than 1/3. AIC: 22,808.9 (intercept only) and 16,456.8 (intercept and covariates).

Table R.A.3. Propensity Score Model: ELA Grade 6

Variable		Value	Est. (SE)
Intercept			-6.900 (0.260)
Grade 5 2016 OP Scale Score*			0.021 (0.030)
Ethnicity	Asian/Pacific Islander		0.089 (0.118)
	Black		-1.200 (0.105)
	Hispanic		-0.846 (0.105)
	Other / Missing		-0.109 (0.127)
English Language Learner			-0.522 (0.200)
Student with Disability			-0.096 (0.090)
Needs/ Resource Category	High Needs		-0.615 (0.063)
	Low Needs		-3.490 (0.266)
	Charter / Religious and Independent		6.287 (1.414)
Joint Management Team Region	Capital District / North Country		2.472 (0.203)
	Long Island		-0.612 (0.367)
	Mid-Hudson		2.158 (0.218)
	Mid-State		3.619 (0.200)
	Mid-West		2.265 (0.206)
	West		1.587 (0.208)
	Other / Missing		0.933 (0.220)
District Minimum Bandwidth	1–9 Gbps		1.017 (0.070)
	>= 10 Gbps		-0.274 (0.104)
	Missing		-4.550 (1.424)
Ratio of CBT Devices to District Enrollment	0.33 - 0.66		0.264 (0.174)
	>= 0.67		1.482 (0.163)
	Missing		n/a

Note. *: Grand-mean centered. Bolded parameter estimates were significant at the $\alpha = .01$ level. Reference students were non-ELL and non-disabled White students attending average needs / resource public schools in the Mid-South JMT region. Their districts reported that the minimum building bandwidth was less than 1 Gbps and a ratio of CBT-eligible devices to enrolled students of less than 1/3. AIC: 19,854.0 (intercept only) and 13,947.2 (intercept and covariates).

Table R.A.4. Propensity Score Model: ELA Grade 7

Variable	Value	Est. (SE)
Intercept		-5.183 (0.213)
Grade 6 2016 OP Scale Score*		0.018 (0.024)
Ethnicity	Asian/Pacific Islander	-0.148 (0.092)
	Black	-0.779 (0.079)
	Hispanic	-0.274 (0.070)
	Other / Missing	-0.135 (0.120)
English Language Learner		-0.406 (0.134)
Student with Disability		-0.317 (0.075)
Needs/ Resource Category	High Needs	-0.081 (0.049)
	Low Needs	-1.256 (0.079)
	Charter / Religious and Independent	6.899 (1.408)
Joint Management Team Region	Capital District / North Country	-0.700 (0.097)
	Long Island	0.069 (0.098)
	Mid-Hudson	-1.401 (0.163)
	Mid-State	1.271 (0.077)
	Mid-West	0.415 (0.083)
	West	-0.927 (0.102)
	Other / Missing	-0.350 (0.092)
District Minimum Bandwidth	1–9 Gbps	0.568 (0.055)
	>= 10 Gbps	0.011 (0.070)
	Missing	-5.663 (1.420)
Ratio of CBT Devices to District Enrollment	0.33 - 0.66	2.040 (0.199)
	>= 0.67	2.230 (0.198)
	Missing	n/a

Note. *: Grand-mean centered. Bolded parameter estimates were significant at the $\alpha = .01$ level. Reference students were non-ELL and non-disabled White students attending average needs / resource public schools in the Mid-South JMT region. Their districts reported that the minimum building bandwidth was less than 1 Gbps and a ratio of CBT-eligible devices to enrolled students of less than 1/3. AIC: 26,857.1 (intercept only) and 20,759.6 (intercept and covariates).

Table R.A.5. Propensity Score Model: ELA Grade 8

Variable	Value	Est. (SE)
Intercept		-6.216 (0.251)
Grade 7 2016 OP Scale Score*		-0.057 (0.030)
Ethnicity	Asian/Pacific Islander	-0.134 (0.097)
	Black	-0.974 (0.104)
	Hispanic	-0.581 (0.096)
	Other / Missing	-0.320 (0.169)
English Language Learner		0.225 (0.155)
Student with Disability		-0.224 (0.091)
Needs/ Resource	High Needs	-0.217 (0.067)
	Low Needs	-0.089 (0.080)
Category	Charter / Religious and Independent	6.199 (1.416)
Joint Management Team Region	Capital District / North Country	0.662 (0.160)
	Long Island	1.210 (0.166)
	Mid-Hudson	1.015 (0.175)
	Mid-State	2.247 (0.147)
	Mid-West	1.606 (0.150)
	West	0.038 (0.170)
	Other / Missing	-0.121 (0.173)
District Minimum Bandwidth	1–9 Gbps	0.439 (0.069)
	>= 10 Gbps	0.464 (0.084)
	Missing	-4.625 (1.428)
Ratio of CBT	0.33 - 0.66	0.579 (0.212)
Devices to District	>= 0.67	1.991 (0.203)
Enrollment	Missing	n/a

Note. *: Grand-mean centered. Bolded parameter estimates were significant at the $\alpha = .01$ level. Reference students were non-ELL and non-disabled White students attending average needs / resource public schools in the Mid-South JMT region. Their districts reported that the minimum building bandwidth was less than 1 Gbps and a ratio of CBT-eligible devices to enrolled students of less than 1/3. AIC: 17,991.3 (intercept only) and 13,815.0 (intercept and covariates).

Table R.A.6. Propensity Score Model: Mathematics Grade 4

Variable	Value	Est. (SE)
Intercept		-5.108 (0.203)
Grade 3 2016 OP Scale Score*		-0.071 (0.034)
Ethnicity	Asian/Pacific Islander	0.256 (0.119)
	Black	-1.162 (0.126)
	Hispanic	-0.802 (0.114)
	Other / Missing	-0.134 (0.138)
English Language Learner		-0.785 (0.227)
Student with Disability		-0.355 (0.114)
Needs/ Resource Category	High Needs	-0.970 (0.083)
	Low Needs	-1.296 (0.120)
	Charter / Religious and Independent	5.985 (1.416)
Joint Management Team Region	Capital District / North Country	1.144 (0.170)
	Long Island	0.614 (0.196)
	Mid-Hudson	1.552 (0.179)
	Mid-State	2.335 (0.162)
	Mid-West	0.263 (0.189)
	West	0.425 (0.178)
	Other / Missing	-0.002 (0.187)
District Minimum Bandwidth	1–9 Gbps	0.407 (0.078)
	>= 10 Gbps	-0.424 (0.125)
	Missing	-5.146 (1.421)
Ratio of CBT Devices to District Enrollment	0.33 - 0.66	-1.704 (0.185)
	>= 0.67	0.834 (0.124)
	Missing	n/a

Note. *: Grand-mean centered. Bolded parameter estimates were significant at the $\alpha = .01$ level. Reference students were non-ELL and non-disabled White students attending average needs / resource public schools in the Mid-South JMT region. Their districts reported that the minimum building bandwidth was less than 1 Gbps and a ratio of CBT-eligible devices to enrolled students of less than 1/3. AIC: 14,700.1 (intercept only) and 11,239.2 (intercept and covariates).

Table R.A.7. Propensity Score Model: Mathematics Grade 5

Variable	Value	Est. (SE)
Intercept		-4.577 (0.180)
Grade 4 2016 OP Scale Score*		-0.061 (0.034)
Ethnicity	Asian/Pacific Islander	0.125 (0.121)
	Black	-1.564 (0.134)
	Hispanic	-0.907 (0.113)
	Other / Missing	0.290 (0.110)
English Language Learner		-1.379 (0.279)
Student with Disability		-0.088 (0.096)
Needs/ Resource	High Needs	0.099 (0.065)
	Low Needs	-1.151 (0.139)
Category	Charter / Religious and Independent	6.743 (1.415)
Joint Management Team Region	Capital District / North Country	0.139 (0.121)
	Long Island	-0.631 (0.172)
	Mid-Hudson	0.608 (0.137)
	Mid-State	1.372 (0.111)
	Mid-West	0.069 (0.126)
	West	-0.454 (0.133)
	Other / Missing	-1.110 (0.149)
District Minimum Bandwidth	1–9 Gbps	0.373 (0.075)
	>= 10 Gbps	0.555 (0.088)
	Missing	-5.521 (1.422)
Ratio of CBT Devices to District Enrollment	0.33 - 0.66	0.226 (0.150)
	>= 0.67	0.618 (0.144)
	Missing	n/a

Note. *: Grand-mean centered. Bolded parameter estimates were significant at the $\alpha = .01$ level. Reference students were non-ELL and non-disabled White students attending average needs / resource public schools in the Mid-South JMT region. Their districts reported that the minimum building bandwidth was less than 1 Gbps and a ratio of CBT-eligible devices to enrolled students of less than 1/3. AIC: 15,841.9 (intercept only) and 12,728.4 (intercept and covariates).

Table R.A.8. Propensity Score Model: Mathematics Grade 6

Variable	Value	Est. (SE)
Intercept		-4.919 (0.196)
Grade 5 2016 OP Scale Score*		0.038 (0.031)
Ethnicity	Asian/Pacific Islander	0.080 (0.120)
	Black	-0.975 (0.113)
	Hispanic	-0.826 (0.120)
	Other / Missing	-0.139 (0.136)
English Language Learner		-0.572 (0.221)
Student with Disability		-0.221 (0.096)
Needs/	High Needs	-1.033 (0.070)
Resource	Low Needs	-2.674 (0.170)
Category	Charter / Religious and Independent	5.733 (1.414)
Joint Management Team Region	Capital District / North Country	0.741 (0.101)
	Long Island	-1.341 (0.198)
	Mid-Hudson	-0.748 (0.183)
	Mid-State	1.676 (0.096)
	Mid-West	0.121 (0.112)
	West	-0.738 (0.123)
	Other / Missing	-0.818 (0.134)
District Minimum Bandwidth	1–9 Gbps	0.984 (0.073)
	>= 10 Gbps	-0.316 (0.108)
	Missing	-4.378 (1.424)
Ratio of CBT	0.33 - 0.66	0.286 (0.176)
Devices to District	>= 0.67	1.500 (0.166)
Enrollment	Missing	n/a

Note. *: Grand-mean centered. Bolded parameter estimates were significant at the $\alpha = .01$ level. Reference students were non-ELL and non-disabled White students attending average needs / resource public schools in the Mid-South JMT region. Their districts reported that the minimum building bandwidth was less than 1 Gbps and a ratio of CBT-eligible devices to enrolled students of less than 1/3. AIC: 19,002.3 (intercept only) and 13,382.0 (intercept and covariates).

Table R.A.9. Propensity Score Model: Mathematics Grade 7

Variable	Value	Est. (SE)
Intercept		-6.135 (0.259)
Grade 6 2016 OP Scale Score*		-0.012 (0.032)
Ethnicity	Asian/Pacific Islander	0.156 (0.106)
	Black	-0.992 (0.104)
	Hispanic	-0.143 (0.089)
	Other / Missing	-0.002 (0.139)
English Language Learner		-1.450 (0.271)
Student with Disability		-0.517 (0.100)
Needs/	High Needs	0.000 (0.061)
Resource	Low Needs	-1.415 (0.106)
Category	Charter / Religious and Independent	6.883 (1.414)
Joint Management Team Region	Capital District / North Country	0.607 (0.135)
	Long Island	0.528 (0.156)
	Mid-Hudson	3.134 (0.219)
	Mid-State	2.233 (0.123)
	Mid-West	1.147 (0.131)
	West	0.030 (0.143)
	Other / Missing	0.347 (0.142)
District Minimum Bandwidth	1–9 Gbps	0.667 (0.069)
	>= 10 Gbps	-0.016 (0.091)
	Missing	-5.520 (1.430)
Ratio of CBT	0.33 - 0.66	1.318 (0.230)
Devices to District	>= 0.67	2.018 (0.226)
Enrollment	Missing	n/a

Note. *: Grand-mean centered. Bolded parameter estimates were significant at the $\alpha = .01$ level. Reference students were non-ELL and non-disabled White students attending average needs / resource public schools in the Mid-South JMT region. Their districts reported that the minimum building bandwidth was less than 1 Gbps and a ratio of CBT-eligible devices to enrolled students of less than 1/3. AIC: 18,500.6 (intercept only) and 14,010.5 (intercept and covariates).

Table R.A.10. Propensity Score Model: Mathematics Grade 8

Variable	Value	Est. (SE)
Intercept		-6.762 (0.356)
Grade 7 2016 OP Scale Score*		-0.057 (0.051)
Ethnicity	Asian/Pacific Islander	0.891 (0.144)
	Black	-1.184 (0.169)
	Hispanic	-0.748 (0.171)
	Other / Missing	-0.177 (0.228)
English Language Learner		-0.560 (0.284)
Student with Disability		0.025 (0.116)
Needs/ Resource Category	High Needs	-1.155 (0.106)
	Low Needs	-3.946 (0.537)
	Charter / Religious and Independent	4.970 (1.419)
Joint Management Team Region	Capital District / North Country	1.999 (0.262)
	Long Island	1.075 (0.357)
	Mid-Hudson	2.245 (0.384)
	Mid-State	3.027 (0.256)
	Mid-West	2.287 (0.262)
	West	0.974 (0.273)
	Other / Missing	-0.128 (0.308)
District Minimum Bandwidth	1–9 Gbps	1.348 (0.124)
	>= 10 Gbps	0.628 (0.147)
	Missing	-3.123 (1.443)
Ratio of CBT Devices to District Enrollment	0.33 - 0.66	0.401 (0.243)
	>= 0.67	1.078 (0.233)
	Missing	n/a

Note. *: Grand-mean centered. Bolded parameter estimates were significant at the $\alpha = .01$ level. Reference students were non-ELL and non-disabled White students attending average needs / resource public schools in the Mid-South JMT region. Their districts reported that the minimum building bandwidth was less than 1 Gbps and a ratio of CBT-eligible devices to enrolled students of less than 1/3. AIC: 8,863.3 (intercept only) and 6,159.0 (intercept and covariates).

Figure R.A.1. Distribution of Propensity Scores Before and After Matching: ELA

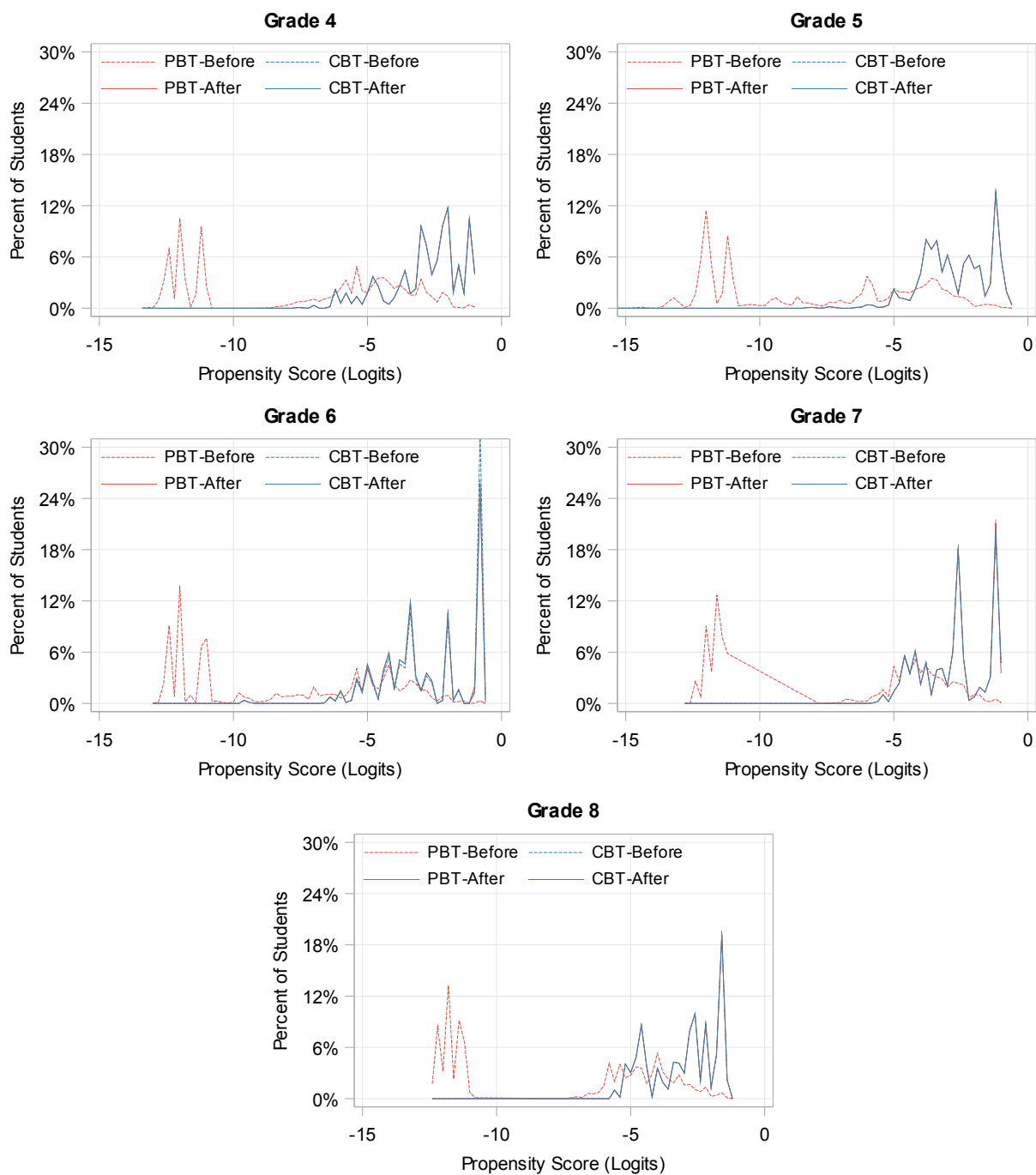


Figure R.A.2. Distribution of Propensity Scores Before and After Matching: Mathematics

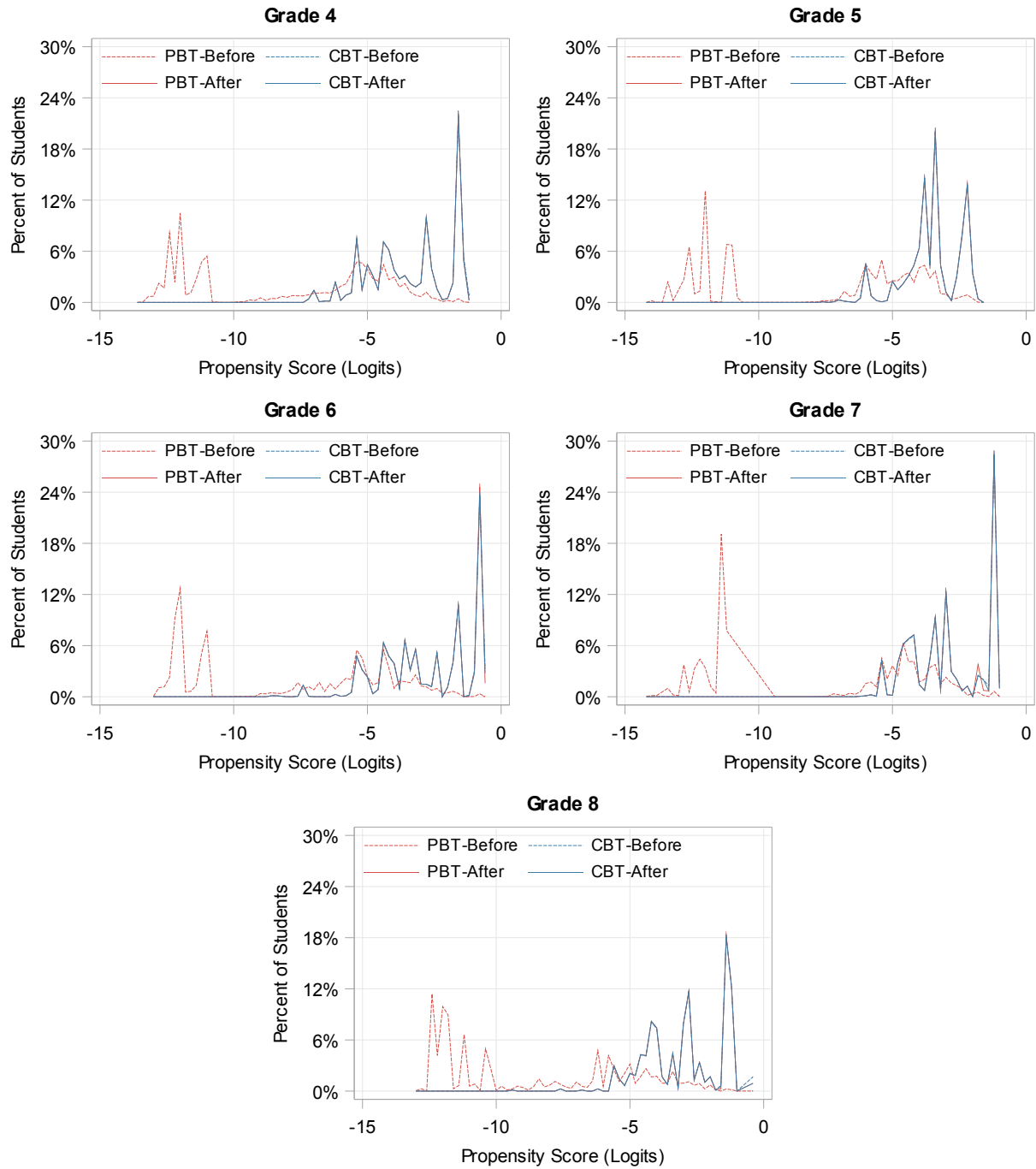


Figure R.A.3. Distribution of Raw Score Differences After Matching: ELA

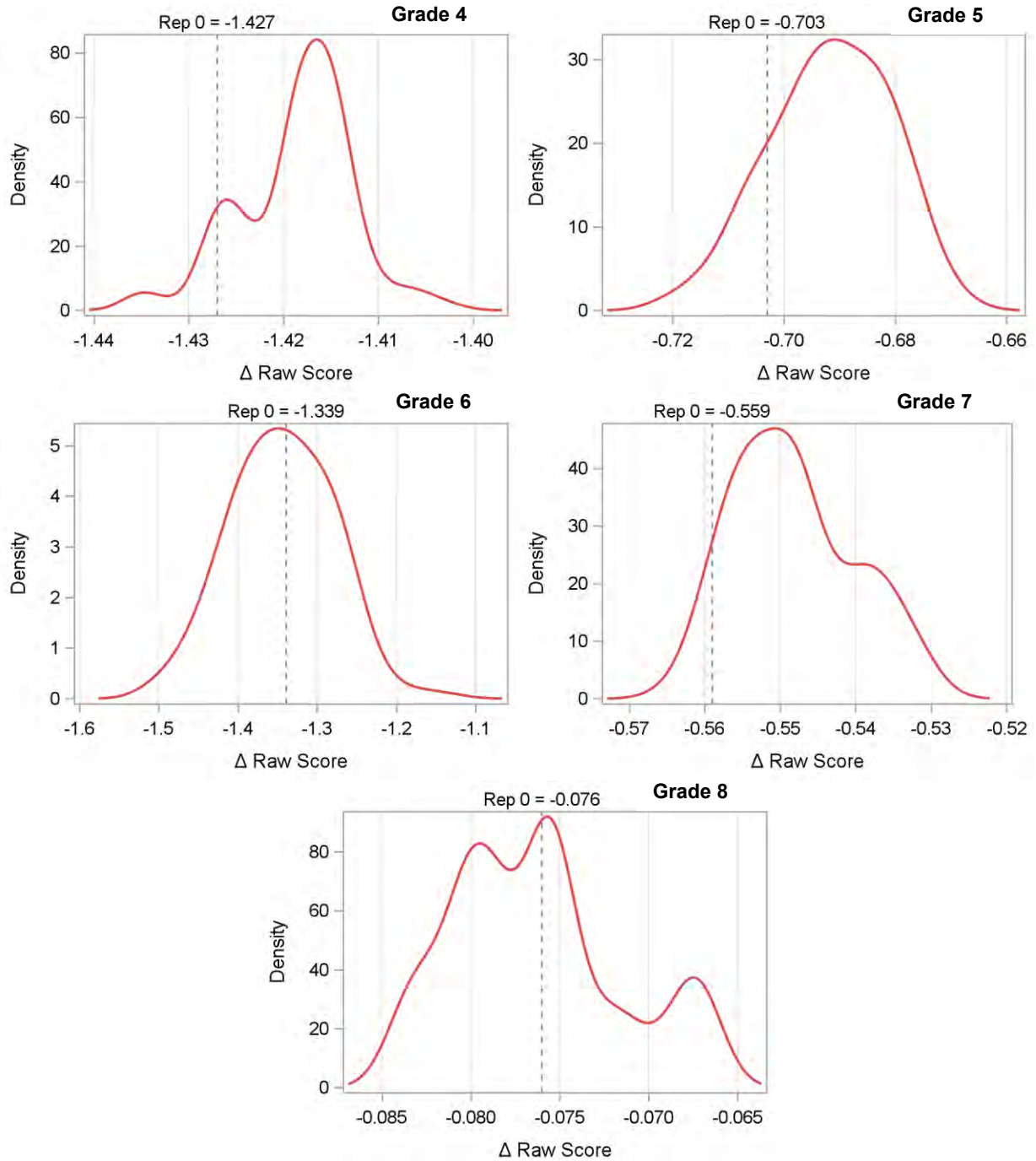
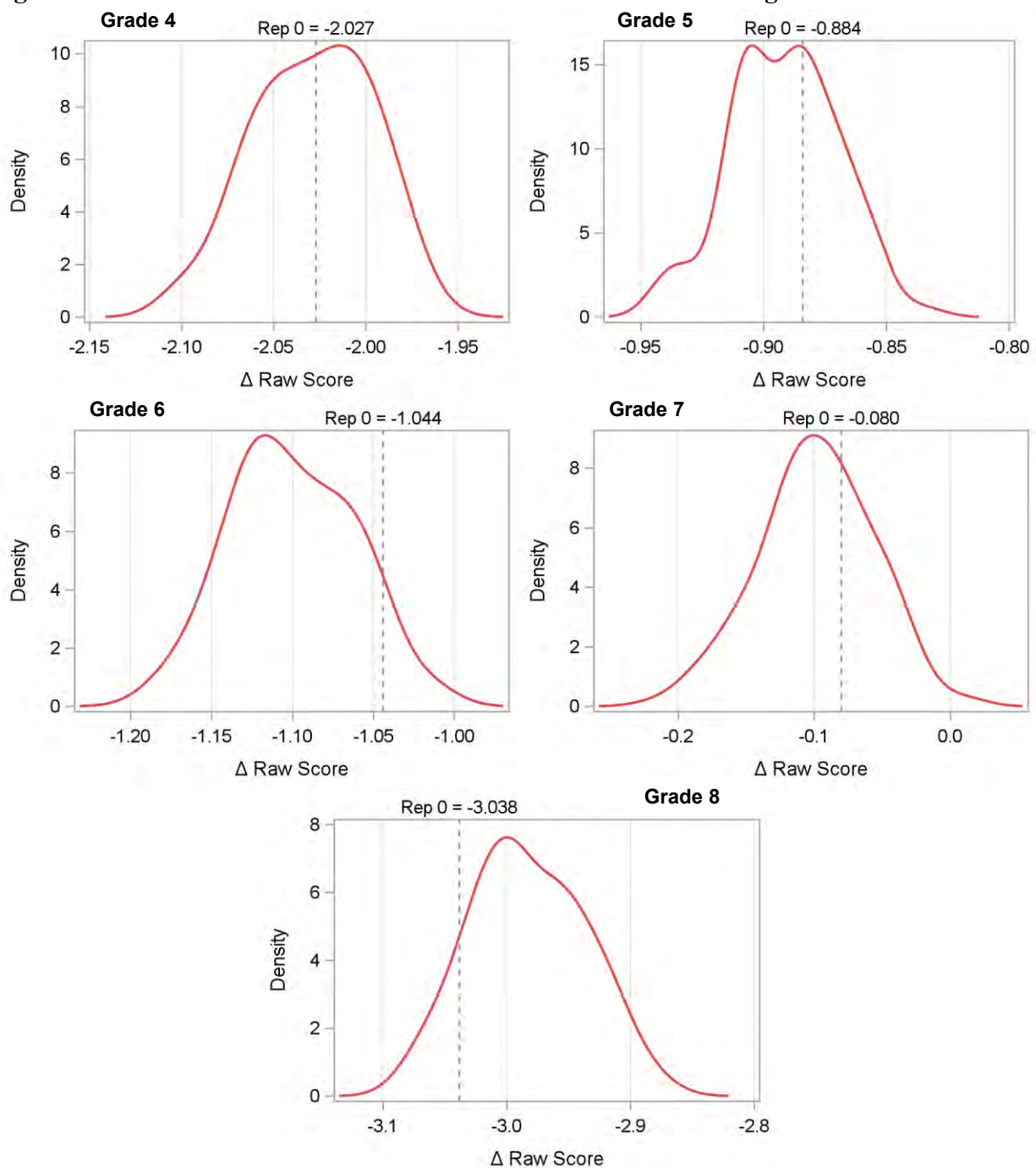


Figure R.A.4. Distribution of Raw Score Differences After Matching: Mathematics



Appendix S: Memo on Operational Test Mode Comparability



THE STATE EDUCATION DEPARTMENT / THE UNIVERSITY OF THE STATE OF NEW YORK / ALBANY, NY 12234

Steven E. Katz
Assistant Commissioner of State Assessment

October 2017

TO: District Superintendents
Superintendents of Public, Religious, and Independent Schools Principals
of Public, Religious, and Independent Schools Charter School Leaders

FROM: Steven E. Katz *Steven E. Katz*

SUBJECT: Comparability of Spring 2017 Grades 3-8 English Language Arts and Mathematics
Paper-based and Computer-based Tests

The purpose of this memorandum is to provide information about the results of the comparability study that was conducted for the Spring 2017 Grades 3-8 English Language Arts (ELA) and Mathematics paper-based and computer-based tests.

Background

In Spring 2017, the Department offered the Grades 3-8 ELA and Mathematics Tests in two administration modes: paper-based testing (PBT) and computer-based testing (CBT). Administering these tests via CBT was optional for schools and those schools that chose to offer CBT made this decision independently for each subject and grade. The Department provided readiness verification tools to help those schools selecting CBT ensure they were well equipped and prepared to provide a successful CBT experience for their students. Additionally, several CBT practice test sessions were made available to CBT schools to familiarize students and teachers with the new CBT delivery system. Each of the CBT practice test sessions featured examples of all types of test questions included on the computer-based tests. This provided the opportunity for students to practice answering both multiple-choice and constructed-response questions on the computer devices they would be using for the actual test.

To further ensure fairness, the Department's contractor, Questar Assessment Inc., conducted a comparability study to identify whether or not there were any differences in student performance that could be attributed to the mode of test administration (i.e., PBT versus CBT). The comparability study methodology and results are summarized below. The findings of this study were used to ensure that students received a score that was representative of their knowledge and skills, regardless of whether they took the tests on paper or computer.

Comparability Study Methodology

Only a small proportion of schools chose to administer the tests via CBT (representing approximately one percent of all test takers). Therefore, the population of students who tested via CBT were not assumed equivalent to the population of students who tested via PBT. In order to select a sample of students who tested via PBT that could be compared to

those students who tested via CBT, a method called propensity score matching was employed. Propensity score matching allowed for the identification of groups of students who tested via PBT that was similar to the groups of students who tested via CBT on a number of school and student characteristics, including achievement on the prior year's test.

Using these characteristics, Questar selected a group of PBT students that matched the group of CBT students for each grade and subject. This allowed for a direct comparison of student results between the two groups. For comparison, the mean scale scores were calculated for each grade and subject by mode of testing. The results are shown in the section below.

Results of Comparability Study

Table 1 shows the scale score means for the PBT and CBT groups on the 2017 English Language Arts Tests by grade as well as the differences in mean scale scores between the matched groups. Table 2 shows these same data for the 2017 Mathematics Tests.

Table 1. *PBT and CBT Means and Differences for Grades 3-8 ELA*

	PBT Scale Score Mean	CBT Scale Score Mean	Difference (Rounded to nearest whole number)
Grade 3	See footnote*		n/a
Grade 4	305.7	300.4	+5
Grade 5	300.5	298.1	+2
Grade 6	300.1	295.6	+5
Grade 7	307.2	305.5	+2
Grade 8	306.1	306.1	0

* Because Grade 3 students have no prior test results on which to match PBT to CBT students, a PBT comparison group was not created and group means were not calculated for this grade level.

Table 2. *PBT and CBT Means and Differences for Grades 3-8 Math*

	PBT Scale Score Mean	CBT Scale Score Mean	Difference (Rounded to nearest whole number)
Grade 3	See footnote*		n/a
Grade 4	309.1	304.1	+5
Grade 5	309.7	307.6	+2
Grade 6	312.5	309.7	+3
Grade 7	311.1	310.8	0
Grade 8	294.3	286.0	+8

* Because Grade 3 students have no prior test results on which to match PBT to CBT students, a PBT comparison group was not created and group means were not calculated for this grade level.

Adjustments to Scores

For those tests in which no difference in mean scale scores between the matched PBT and CBT groups was observed, no adjustment was made to any students' scale scores. For those tests in which a difference in mean scale scores between the two comparable groups was observed, the scale scores for all students who took the test in that grade via CBT, (which was the lower scoring mode in all such instances during this administration), were adjusted by adding the number of scale score points shown in the "Difference" columns of Tables 1 and 2 to the CBT students' scale scores, up to the maximum attainable scale score. Thus, the scale score adjustments for students who tested via CBT, shown in Table 3 below, reflect the differences between the PBT and CBT groups found in the comparability study. These slight adjustments ensured that students who demonstrated comparable proficiencies in their knowledge and skills received comparable scores whether they tested on paper or on computer.

Table 3. *Summary of Scale Score Adjustments for CBT*

	ELA Scale Score Adjustment	Math Scale Score Adjustment
Grade 3	+4*	+4*
Grade 4	+5	+5
Grade 5	+2	+2
Grade 6	+5	+3
Grade 7	+2	0
Grade 8	0	+8

* Because Grade 3 students have no prior test results on which to match PBT to CBT students, a PBT comparison group was not created and group means were not calculated for this grade level. Instead, the mean adjustment for the other elementary grades for which a comparison was possible (i.e., Grades 4 & 5) was applied to the scores of Grade 3 students who tested via CBT.

For questions concerning the Grades 3-8 ELA or Mathematics Tests, please email the [Office of State Assessment](#) or call 518-474-5902. For questions concerning CBT, please email [CBT support](#).